Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN
INGEGNERIA ELETTRONICA, INFORMATICA
E DELLE TELECOMUNICAZIONI
Ciclo XXIV

Settore Concorsuale di afferenza: 09/E3 ELETTRONICA
Settore Scientifico-disciplinare: ING-INF/01 ELETTRONICA

NUMERICAL MODELLING OF GRAPHENE NANORIBBON-FETS FOR ANALOG AND DIGITAL APPLICATIONS

Presentata da: Ilaria Imperiale

Coordinatore Dottorato:
Chiar.mo Prof. Ing.
Luca Benini

Relatore:
Chiar.mo Prof. Ing.
Massimo Rudan

Esame finale anno 2012

Sommario

Il grafene, uno strato monodimensionale di atomi di carbonio disposti a nido d'ape, è stato recentemente isolato a partire dalla grafite. Questo materiale dispone di proprietà fisiche molto interessanti, tra cui eccellenti mobilità elettroniche, capacità di trasporto di corrente e conduttività termica. Inoltre, i portatori si muovono all'interno del grafene in condizioni quasi balistiche, e la sua struttura planare ed il suo spessore pari ad uno strato atomico fanno supporre che transistor ad effetto di campo (field-effect transistors, FETs) che utilizzano il grafene come materiale di canale sarebbero poco condizionati dagli effetti di canale corto, e che l'integrazione e lo *scaling* del grafene sarebbero più semplici di quelli di altri materiali emergenti per le applicazioni post-CMOS. In considerazione di questo, il grafene negli ultimi tempi è stato oggetto di grande studio, come potenziale candidato per l'utilizzo in dispositivi su scala nanometrica per applicazioni elettroniche. Il maggiore limite per l'utilizzo del grafene nei dispositivi elettronici è il fatto che è un semi-metallo, ossia non dispone di un gap energetico. La presenza di un gap energetico è essenziale nei transistor digitali, che necessitano di un gap per chiudere il canale conduttivo quando il dispositivo è nello stato di OFF. In questa tesi sono stati studiati i graphene nanoribbons (GNRs), che sono delle sottili strisce di grafene, nei quali un gap energetico è causato dal confinamento quantistico delle cariche nella direzione trasversale.
Dato che i GNR-FETs realizzati sperimentalmente sono ancora lontani dallo essere ideali, essenzialmente per l'eccessiva larghezza e per la rugosità ai bordi, una descrizione accurata dei fenomeni fisici presenti in questi dispositivi è necessaria per fare valutazioni corrette riguardo alla performance di queste nuove strutture. Con questo fine, un codice è stato sviluppato e utilizzato per studiare la performance di GNR-FETs di larghezza da 1 a 15 nm. Data l'importanza di una descrizione accurata degli effetti quantistici nel funzionamento dei dispositivi in grafene, è stato utilizzato un modello di trasporto completamente quantistico: la dinamica degli elettroni è stata descritta attraverso un modello di Hamiltoniano tight-binding (TB) e il trasporto è stato risolto con il formalismo delle funzioni di non equilibrio di Green (NEGF). Sono stati considerati sia il trasporto

di tipo balistico che dissipativo. L'interazione elettrone-fonone è stata inclusa nell'approssimazione auto-consistente di Born.
In considerazione della diversa ampiezza del gap energetico di cui dispongono, i GNR stretti sono potenziali candidati per le applicazioni digitali, mentre quelli di larghezza maggiore per le applicazioni a radiofrequenza.

Abstract

Graphene, that is a monolayer of carbon atoms arranged in a honeycomb lattice, has been isolated only recently from graphite. This material shows very attractive physical properties, like superior carrier mobility, current carrying capability and thermal conductivity. In addition, the carriers move inside graphene in quasi-ballistic conditions and its planar structure and atomic thickness suggest that field-effect transistors (FETs) made of graphene as channel material would be slightly affected by short-channel effects and its integration and scaling could be easier than that of other emerging material for post-CMOS applications. In consideration of that, graphene has been the object of large investigation as a promising candidate to be used in nanometer-scale devices for electronic applications. The main disadvantage for the application of graphene in electronic devices is the fact that it is a semi-metal, namely it does not show any energy band-gap. The presence of a band-gap is essential in digital transistors, that require a band-gap to close the conductive channel when the device is in the OFF state. In this work, graphene nanoribbons (GNRs), that are narrow strips of graphene, for which a band-gap is induced by the quantum confinement of carriers in the transverse direction, have been studied.
As experimental GNR-FETs are still far from being ideal, mainly due to the large width and edge roughness, an accurate description of the physical phenomena occurring in these devices is required to have valuable predictions about the performance of these novel structures. A code has been developed to this purpose and used to investigate the performance of 1 to 15-nm wide GNR-FETs. Due to the importance of an accurate description of the quantum effects in the operation of graphene devices, a full-quantum transport model has been adopted: the electron dynamics has been described by a tight-binding (TB) Hamiltonian model and transport has been solved within the formalism of the non-equilibrium Green's functions (NEGF). Both ballistic and dissipative transport are considered. The inclusion of the electron-phonon interaction has been taken into account in the self-consistent Born approximation.
In consideration of their different energy band-gap, narrow GNRs are expected to be suitable for logic applications, while wider ones could be

promising candidates as channel material for radio-frequency applications.

Contents

Introduction

Current integrated-circuit (IC) technology is driven by the silicon transistor, or more importantly, the ability to increasingly scale down the transistor size in order to enhance the performance of an individual transistor while also increasing the total number of transistors for a given area. However, as feature sizes of silicon transistors approach the nanometer scale, transistor performance no longer scales in proportion with device dimensions, particularly channel length [1]. In recent years, the request for increasing performance and reducing area occupancy for both active and passive electronic components has pushed the scaling process every day closer to the physical limits of the silicon-based MOS technology. Thus, the research has been oriented towards the investigation of new nanoscale devices capable to overcome the main limitations observed for silicon MOSFETs at the nanometer scale, like short channel effects (SCEs) and gate leakage. In particular, both the introduction of new architectures and new channel materials have been proposed. Among those, carbon-based materials like graphene and carbon nanotubes have attracted large interest in the scientific community, due to their fascinating electrical properties. Graphene is a monolayer crystal of carbon atoms arranged in a hexagonal structure of carbon atoms arranged in a honeycomb lattice. It is the fundamental building block of graphitic materials, and thus is important in determining the electronic properties of other carbon allotropes such as graphite (3-D stack of graphene sheets), carbon nanotubes (1-D rolled up graphene cylinder), and fullerenes (0-D molecules of wrapped-up graphene with the introduction of pentagons), shown in fig. 1. In fact, graphite can be viewed as a stack of weakly bonded graphene layers and carbon nanotubes (CNTs) can be considered as resulting from the folding of a graphene sheet to a cylinder. Although graphene properties have been known for a long time, only in 2004 a research group at Manchester University succeded in isolating graphene from graphite [2]. The employed technique was based on a repetitive exfoliation of a graphitic block using adhesive tape and the subsequent deposition of the flakes onto an oxidized silicon wafer. Although it was not obtained a perfect single layer of graphene, that was the first reported experiment in which graphene

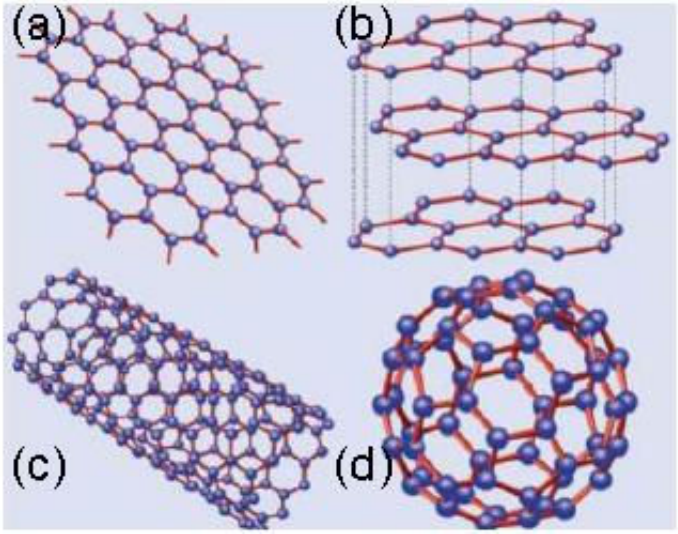

Figure 1: Allotropes of carbon include: a) graphene, b) graphite, c) CNTs and d) fullerenes.

layers had been isolated. Since then, the fabrication of graphene has significantly improved and rapid advancements have been achieved in the understanding of the physical mechanisms typical of this material. The excellent electrical and thermal properties of graphene hold great promise for applications in future IC technology [3]. The interesting properties of graphene arise from its two-dimensional structure that confines electrons in one atomic layer and that causes charge carriers to behave as massless Dirac fermions [4, 5] and to its low density of states (DOS) near the Dirac point, which causes the Fermi energy to shift significantly with variation of carrier density [6]. Early measurements suggested graphene has high intrinsic mobility ($\approx$ 40000 cm^2/Vs) [7] and high thermal conductivity ($\approx$ 600 W/m K) [8], both significantly greater than in silicon and other standard semiconductors. Therefore, graphene has attracted large attention as a promising candidate for future electronics, with specific interest for radiofrequency applications. A graphene transistor fabricated from graphene epitaxially formed on a SiC wafer has demonstrated a cutoff frequency as high as 100 GHz [9]. In addition, optical properties of graphene make it a promising material for IR optoelectronics [6] and its high electrical and thermical conductivity, makes it an attractive material for future high-speed interconnects. The main drawback for the application of graphene in electronic devices is the fact that it is a semi-metal, namely it does not present an energy band-gap, as shown in 2. The presence of a band-gap is essential in digital transistors, that require a band-gap to close the conductive channel when the device is in the OFF state. Therefore, the lack of a band gap results in very poor I_{ON}/I_{OFF} ratio, too low to be used

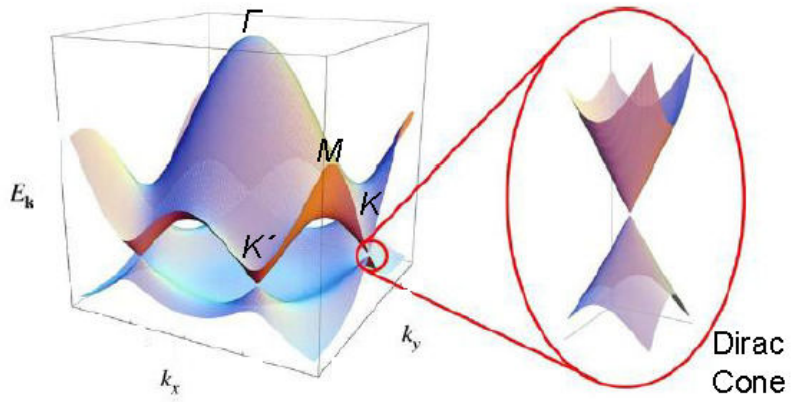

Figure 2: Band structure of graphene showing six Dirac (K and K) points and a linear dispersion relationship around them. Image taken from [11].

in digital logic [10]. Recently, several ways have been proposed in the literature to create a band gap in graphene. One of the most interesting techniques is patterning graphene into narrow ribbons, so that the carriers are confined in a quasi 1-dimensional layer, this resulting in the opening of an energy band-gap. With respect to carbon nanotube field-effect transistors (CNT-FETs), FETs that use graphene nanoribbons as the channel material (GNR-FETs) exhibit comparable performance, reduced sensitivity on the variability of channel chirality, and similar leakage problems due to band-to-band tunneling [12]. However, as carbon nanotubes [13, 14], graphene nanoribbons can be either seminconducting or metallic. A sketch of the two main edge orientations are depicted in fig.3. Zigzag GNRs are metallic, while armchair can be either semiconducting or metallic, depending on the number of atoms in the lateral cross-section. This property is very attractive, since it could lead to the introduction of a fully carbon-based tecnology, where both the active devices and interconnects could be made of the same material, thus avoiding technology compatibility issues. As far as the energy band-gap width is concerned, large theoretical and experimental work has been carried out in recent years to investigate the energy band-gap engineering in graphene nanostructures. For example, in [16], from the analysis of experimental data carried out by the research group guided by Prof. Kim at Columbia University it was shown that

$$E_{gap} = \alpha(W - W) \tag{1}$$

where α could range between 0.2 - 1.5 eV nm and $W* \approx W_0$, where W_0 is given by the consideration that $W - W_0$ is the active GNR width participating in charge transport. Figure 4, presented in [16] is reported here for

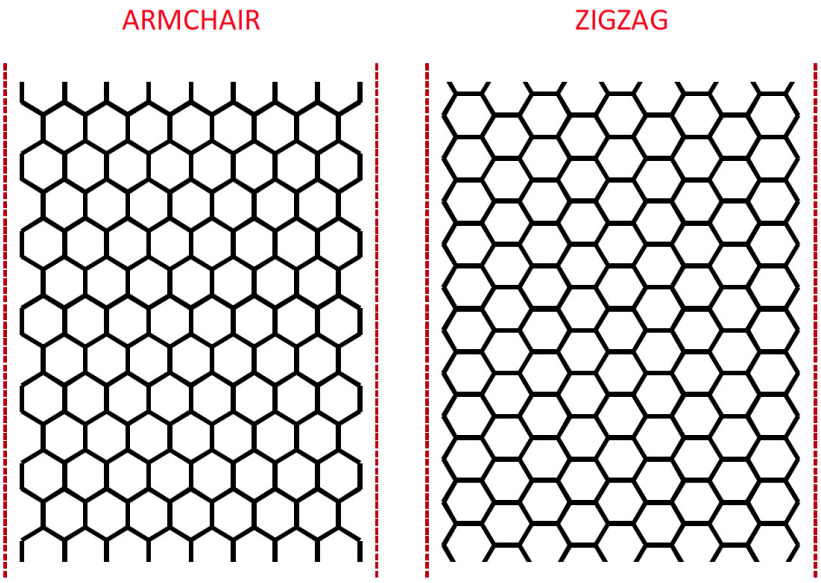

Figure 3: Main orientations of graphene nanoribbons. Image taken from [15].

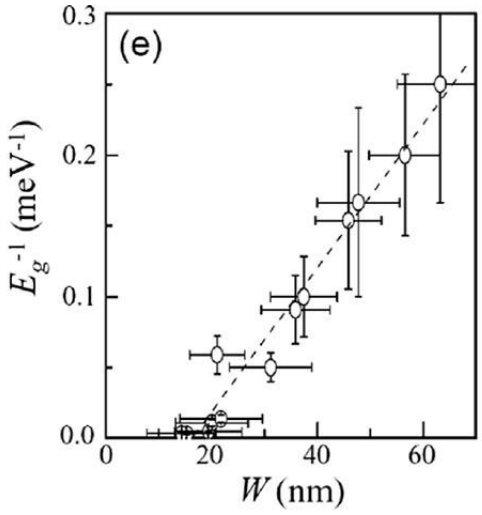

Figure 4: Experimental data of 1/energy band-gap vs. GNR width shown in [16].

convenience. Practically, the possibility to obtain a tunable energy band-gap for graphene nanoribbons by lithographic process was demonstrated. Therefore, graphene nanoribbons have been regarded as a material that potentially offers large design flexibility.
This work will be focused on the investigation of electrical properties of field-effect transistors that use graphene nanoribbons as channel material (GNR-FETs): in chapter. 1, the model used to simulate armchair GNR-FETs will be discussed. In chapter 2 will the simulation results regarding

very narrow devices mainly for digital applications and a physical insight of the main phenomena that affect their performance will be presented; in chapter 3 the simulation study will be extended to wider GNR-FETs, that are promising candidates for high-frequency applications, and design consideration to maximise the performance will be suggested.

Chapter 1

Model

The combination of the tight-binding formulation to describe the Hamiltonian and the NEGF formalism to address the transport issue provides the state-of-the-art technique to model quantum transport in carbon-based materials. Since the TB uses atomic orbitals as basis functions, it can potentially describe the real atomic structure of the material, therefore including the effect of atomistic defects, such as rough edges in GNRs. However, complete physical insight is generally achieved at the expense of long computational times, which are not practical for device optimization studies. Thus, in this work a simplified but accurate approach is presented for the simulation of transport in armchair GNR devices. The idea behind this approach, called mode-space tight-binding, is that the graphene nanoribbon behaves as a confined structure in the transverse direction. In fact, due to structural confinement, the two dimensional-graphene dispersion relation splits up in many 1D subbands, whose separation in energy is inversely proportional to the GNR width. If the subband index is a good quantum number, namely if the electrons travel through the device without changing subband, a large computational advantage can be achieved by considering a separate transport problem for each subband and by simulating only those subbands that are present within the energy interval under investigation. The NEGF formalism can be seen as the quantum analogue of the Boltzmann equation. While Boltzmann's equation combines Newton's law with a statistical description of interactions, the NEGF formalism combines quantum dynamics with an analogous description of interactions. The complete formulation of NEGF can be found in [18]. This chapter is organized as follows: in section 1, the adopted TB model for GNRs will be introduced and the general procedure for solving the NEGF equations in real space will be presented. In sec. 2, the models for the inclusion of electron-phonon scattering, both acoustic and optical will be illustrated. Finally, in sec. 3 some considerations regarding the self-consistency of the solution will be provided.

1.1 Overview of the model

An armchair GNR has been used as channel material of a FET. The nanoribbon is generally sandwiched between two oxide layers and the electrostatic potential over it is modulated by the field-effect of two short-circuited gate contacts. The source and drain ends of the GNR are assumed to be doped and to be connected to two semi-infinite leads, made of the same GNR as the device region. The two leads are conceptually supposed to be connected to two large contacts that maintain them in equilibrium. Therefore, the particles injected from each of the two leads into the device can be described by an equilibrium Fermi distribution. However, the Fermi levels of the source and drain leads, E_{FS} and E_{FD}, are in general different, and their difference is equal to -qV_{DS}, where q is the electronic charge and V_{DS} the applied voltage between the drain and the source. The purpose of the simulation is to compute the current I_{DS} that flows from source to drain as a function of the applied voltages V_{DS} and V_{GS}, in a steady-state condition. Charge transport is assumed to occur only inside the GNR, which can be considered to be a periodic structure along the longitudinal direction. For an armchair ribbon, the unit cell or slab is made of two rows of dimers. Its length is equal to $3\ a_{CC}$ where a_{CC} = 1.42 Åis the carbon interatomic distance. For reference, a slab taken from an N_A = 13 armchair GNR, with N_A is the number of dimers in the slab, is illustrated in fig.1.1. The TB Hamiltonian is introduced to quantum-mechanically describe the electron dynamics inside the GNR. A a set of orthogonal p_z orbitals, one for each carbon atom, is in general sufficient to describe transport in graphene-related materials. Indicating with $|l, \alpha\rangle$ the orbital associated with the atom α within the slab l, the generic matrix element of the Hamiltonian $H_{l\alpha,m\beta}$ is written as

$$\langle l, \alpha | H | m, \beta \rangle = t_{l\alpha,m\beta} + \delta_{l\alpha,m\beta}\, U_{l\alpha} \tag{1.1}$$

where $\delta_{l\alpha,m\beta}$ is the Kronecker delta and $U_{l\alpha}$ is the electrostatic potential energy at the (l; α) atom site. The parameter $t_{l\alpha,m\beta\rangle}$ is usually taken as a

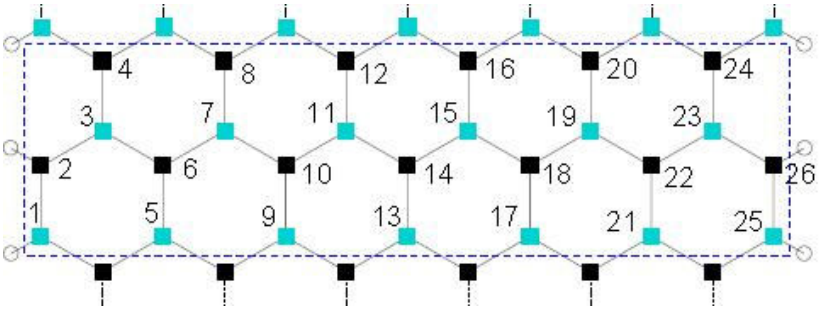

Figure 1.1: 1D elementary cell or slab of an armchair GNR with N_A = 13.

fitting parameter with respect to DFT: for graphene, one can obtain an accurate model by setting the $t_{|l\alpha,m\beta\rangle}$ = t1 if the atoms (l; α) and (m; β) are first nearest neighbors, and equal to zero otherwise. In this work the approach proposed by [17] has been followed: t1 has been taken equal to -2.7 eV for the internal atom pairs and modified with a factor δ equal to 0.12 eV for the atom pairs along the edges of the GNR, to describe the passivation of the edges by hydrogen atoms. In the real space (RS) approach, the transport problem is formulated within the NEGF formalism [18, 19] using the Hamiltonian described above. The retarded Green's function G^r at the energy E is defined by

$$[(E + i\eta)I - H]G^r = I \tag{1.2}$$

with η an infinitesimal positive quantity. For convenience, the quantity A will be defined, with $A = (E + i\eta)I - H$. The matrix equation 1.2 is of infinite dimension because it describes the entire structure made of the device region plus the two semi-infinite source and drain leads. It can be proved [18] that if one can solve the problem in the leads, namely

$$A_S \; g_S^r = I_S \tag{1.3}$$

$$A_D \; g_D^r = I_D \tag{1.4}$$

then it is possible to define two self-energies Σ_S^r and Σ_D^r:

$$\Sigma_S^r = A_{CS} \; g_S^r \; A_{SC} \tag{1.5}$$

$$\Sigma_D^r = A_{CD} \; g_D^r \; A_{DC} \tag{1.6}$$

where A_{CS} defines the matrix that couples the source and the channel. Similar expressions hold for A_{SC}, A_{CD}, A_{DC}. The problem in the device region therefore becomes

$$(A_C - \Sigma_S^r - \Sigma_D^r) \; G_C^r = I_C \tag{1.7}$$

Indicating G_C^r with G^r and explicitating the dependence on energy, eq.1.8 takes the form:

$$[(E + i\eta)I - H_C - \Sigma_S^r(E) - \Sigma_D^r(E)]G^r(E) = I \tag{1.8}$$

Assuming that the electrostatic potential in the first and last slabs of the device region is replicated periodically in each slab of the semi-infinite source/drain lead, the self-energies Σ_S^r and Σ_D^r can be numerically computed using the iterative algorithm proposed in [20]. In the simulations presented in this thesis, the convergence factor has been set equal to zero inside the device region (in order to guarantee current conservation) and

to 10^{-7} eV in the leads.
The electron/hole correlation functions are given by

$$G^{</>}(E) = G^r(E)[\Sigma_S^{</>}(E) + \Sigma_D^{</>}(E)]G^a(E) \tag{1.9}$$

where $G^a = G^{r\dagger}$ is the advanced Green's function († represents the Hermitian-transpose operator).
The self-energies $\Sigma_S^{</>}(E)$ describe the in-scattering of electrons/holes from the source lead into the device region and, according to the previously mentioned hypothesis of thermalized contacts, are given by

$$\Sigma_S^{<}(E) = i\ \Gamma_S(E)\ f_S(E) \tag{1.10}$$

$$\Sigma_S^{>}(E) = -i\ \Gamma_S(E)\ [1 - f_S(E)] \tag{1.11}$$

where $\Gamma_S = \Sigma^r - \Sigma^a$ is the broadening function and

$$f_S(E) = \frac{1}{exp(\frac{E-E_{FS}}{k_B T} + 1)} \tag{1.12}$$

is the the Fermi function of the source lead and k_B is the Boltzmann constant. Analogous expressions can be used for the drain lead, with appropriate substitutions. A common temperature T is assumed for both contacts. From eq.1.9, one can calculate the electron and hole numbers at the (l; α) atom site as

$$n_{l\alpha} = -2i \int_{E_i(l,\alpha)}^{\infty} \frac{1}{2\pi}\ G^{<}(l,\alpha;l,\alpha;E)\ dE \tag{1.13}$$

$$p_{l\alpha} = 2i \int_{-\infty}^{E_i(l,\alpha)} \frac{1}{2\pi}\ G^{>}(l,\alpha;l,\alpha;E)\ dE \tag{1.14}$$

where $\mathrm{E}_i(l;\alpha)$ is the intrinsic Fermi level, assumed equal to the potential energy $\mathrm{U}_{l\alpha}^*$, and the factor of 2 is due to spin degeneracy. Finally, the current is calculated as

$$I_{DS} = \frac{2q}{h} \int_{-\infty}^{\infty} 2\Re\{Tr[H(l,l+1)G^{<}(l+1;l;E)]\}\ dE \tag{1.15}$$

where h is the Planck constant and symbols $\Re$ and Tr indicate the real part and the trace on the orbital index, respectively. Since only coherent transport is considered here, it can be proved that eq.1.15 is also equivalent to

$$I_{DS} = \frac{2q}{h} \int_{-\infty}^{\infty} T(E)[f_S(E) - f_D(E)]\ dE \tag{1.16}$$

where T(E) is the transmission function of the Landauer formalism [18]

$$T(E) = Tr[\Gamma_S G^r \Gamma_D G^a] \tag{1.17}$$

It is worth observing that, since H is a block-tridiagonal matrix, with each block representing the coupling between two adjacent slabs, the only non-null block of Σ_S^r is the first term $\Sigma_S^r(1,1)$, while the only non-null block of Σ_D^r is the last term $\Sigma_D^r(N,N)$, with N the number of slabs in the simulation domain. As a consequence, by directly expanding 1.13, 1.14 and 1.16, it can be obtained that the only blocks of G^r which are needed to compute charge and current are those related to the first and last columns, i.e. $G_{i,1}^r$ and $G_{i,N}^r$ with i = 1...N. Therefore, a recursive algorithm can be used to compute only those blocks [19].
From the computational point of view the real-space tight-binding is heavy, since it involves the calculation of the elements and the inversion of square matrices whose size is equal to the number of orbitals in each slab. Since the model proposed here is based on a single p_z orbital approximation, the dimension of the matrices turns out to be equal to the number of atoms in each slab. For example, for a GNR with smooth edges with N_A=13, the dimension of the generic matrix is 26. It is apparent that this approach is not suitable for the simulation of wide GNRs. Since the purpose of the work is to perform a simulation study on GNR-FETs with width up to 15 nm (corresponding to N_A= 124), this poses the need of finding an approximate but accurate model to limit the use of real space method. As similar methods in literature, mode-space tight-binding is characterized by a change of representation from real space (RS), where the unknown quantities and the Hamiltonian are expressed in terms of atomic orbitals, to mode-space (MS), where the basis is composed of a convenient subset of the transverse eigenvectors (modes).
Given a unitary matrix V, 1.8 in the real space can be transformed into an MS equation of the type:

$$[(E + i0^+)I - \widetilde{H_C} - \widetilde{\Sigma^r}_S(E) - \widetilde{\Sigma^r}_D(E)]\widetilde{G^r}(E) = I \tag{1.18}$$

where the quantities indicated by ˜ identify the representations in the mode-space of the corresponding quantities in the real space. For example,

$$\widetilde{H_C} = V^\dagger H_C V \tag{1.19}$$

$$\widetilde{G^r}(E) = V^\dagger G^r(E) V \tag{1.20}$$

Similar expressions hold for $\widetilde{\Sigma_S^r}$(E) and $\widetilde{\Sigma_D^r}$(E).
Once $\widetilde{G^r}$(E) is known, the RS solution can be easily reconstructed by inverting eq.1.20. The benefit of the mode-space approach is that solving

1.18 instead of 1.8 is computationally advantageous if $\widetilde{H_C}$ can be written as a block diagonal matrix apart from an index reordering, thus giving rise to an independent problem for each block (mode decoupling). An additional simplification is achieved if only a subset of these independent problems gives a significant contribution to G^r in the simulated energy window, thus allowing one to neglect the other blocks (mode truncation). Clearly, the efficiency of the MS method depends on the degree with which these two simplifications can be accurately performed. Thus, the selection of the modes to be retained in the calculations and the identification of the coupled and uncoupled modes play a crucial role in the MS approach. Here the transformation matrix V has been chosen as a block diagonal matrix, which has in the columns of its block of index l the orthonormal eigenvectors at k = 0 (modes) of the slab l, computed with the electrostatic potential made periodic along the longitudinal direction. Regarding the mode coupling of an ideal armchair GNR with uniform electrostatic potential (the Hamiltonian of which is periodic), it can be studied by comparing the band structure of the RS Hamiltonian with that of the MS Hamiltonian obtained with a specific mode selection, i.e. using a specific subset of the eigenvectors at k = 0 (group of modes) as columns of the generic diagonal block of the transformation matrix. If the selected modes are sufficient to accurately reproduce the desired portion of the RS band structure, it means that it is reasonable to consider them uncoupled from the others. In this work, the approach proposed by [21] has been followed: first the modes are split into several groups and a mode is considered to be coupled only with the other modes within the same group but not with the ones belonging to different groups (decoupling criterion). Secondly, only the groups containing at least one of the N_b lowest energy conduction modes or one of the N_b highest energy valence modes are retained (truncation criterion), where N_b is the number of conduction/valence band pairs that are required to be computed with sufficient accuracy. The algorithm is then used for the selection of modes prior to the simulation of devices with regular GNRs, namely GNRs made of the periodic repetition of an elementary slab. Since the presence of a non-uniform potential along the axis of a regular GNR does not represent a serious cause of mode coupling, selection criteria based essentially on the observation of the eigenvalues with constant potential are in general sufficient [22]. Remarkably, although the formulation of the model has been presented energy per energy, the proposed method is applicable to the case of incoherent scattering as well.

1.2 Inclusion of electron-phonon interaction

The electron-phonon interaction is included within a perturbative model within the self-consistent Born approximation. Thus, the phonon system is considered unperturbed by the interactions with the electron gas, therefore the self-energy induced by the presence of phonon scattering can be expressed by:

$$\Sigma_{ph}^{</>} = G^{</>}\, D^{</>} \tag{1.21}$$

with $D^{</>}$ the less-than and greater-than Green's functions of the unperturbed phonon bath. In eq. 1.21, real space is assumed and the explicit dependence on energy is omitted. Details on the explicit calculation of $\Sigma_{ph}^{</>}$ can be found in [18]. Remarkably, the solution of the kinetic equations requires also the knowledge of the retarded self-energy Σ^r, that can be calculated from the relation [23]:

$$\Sigma^r(E) = P \int \frac{1}{2\pi} \frac{\Gamma(\epsilon)}{E-\epsilon} - i\frac{\Gamma(E)}{2}\, d\epsilon \tag{1.22}$$

where P is a principal value integral on the complex plane, and Γ is defined as

$$\Gamma(E) = i[\Sigma^{>}(E) - \Sigma^{<}(E)] \tag{1.23}$$

The real part of Σ^r, represented by the first term of the right side of 1.22, is a non-hermitian energy contribution giving a shift of the particle energy levels, while the second term is associated to the the scattering rate due to the electron-phonon interaction. Therefore, the evaluation of 1.22 can be performed analytically only for the case of elastic scattering while requiring a numerical evaluation of the principal value integral in the case of inelastic scattering process. Since this calculation can be computationally expensive due the necessity of the simultaneous knowledge of the Green's functions for any energy, its contribution is generally omitted. The impact of this assumption has been investigated in [24]. Thus, generally eq.1.22 is well approximated with

$$\Sigma^r(E) \approx -i\frac{\Gamma(E)}{2} \tag{1.24}$$

It is important to notice that the relations above introduce a dependence of Σ^r on $G^{<}$. This implies, that in presence of the electron-phonon interaction eq. 1.8 and

$$G^{</>} = G^r \Sigma^{</>} G^{r\dagger} \tag{1.25}$$

are coupled through a non linear relation. While in the ballistic case those two equations separately describe the dynamics and the statistical properties of the system, when the phonon scattering is included, a self-consistent iterative solution of the two equations with the phonon self-energy functions is required.

1.2.1 Acoustic phonon scattering

As far as the scattering with acoustic phonons is concerned, the term $\Sigma^r_{ph} = \Sigma^r_{AP}$ reads, in the real space approach [18]:

$$\Sigma^r_{AP}(i\alpha, j\beta) = K_{AP}\, G^r(i\alpha, j\beta)\delta_{ij}\delta_{\alpha\beta} \tag{1.26}$$

where K_{AP} is given by

$$K_{AP} = \frac{D_{AP} k_B T}{m_c v_s^2} \tag{1.27}$$

in which D_{AP} is the deformation potential, k_B is the Boltzmann constant, T is the absolute temperature, m_c is the carbon atom mass and v_s is the sound velocity in carbon. The phonon parameters for the 2D graphene longitudinal acoustic mode are employed [25], namely D_{AP} = 16 eV and v_s approximated to $2 \cdot 10^4$ m s^{-1}. According to the change of basis set in the mode space $\widetilde{\Sigma^r_{AP}}$ can be obtained as:

$$\Sigma^{r'}_{AP} = V^\dagger \Sigma^r_{AP} V \tag{1.28}$$

Therefore,

$$\widetilde{\Sigma^{r'}_{AP}}(i\alpha, j\beta) = \sum_{l\gamma, m\delta} [V^\dagger_{i\alpha, l\gamma} \Sigma^r_{AP}(l\gamma, m\delta) V_{m\delta, j\beta}] \tag{1.29}$$

thus

$$\Sigma^{r'}_{AP}(i\alpha, j\beta) = K_{AP} \sum_{l,\gamma} [V^\dagger_{i\alpha, l\gamma} G^r(l\gamma, l\gamma) V_{l\gamma, j\beta}] \tag{1.30}$$

Since V is a blockdiagonal matrix, it holds:

$$V_{i\alpha, l\gamma} = \delta_{i,l} V_i(\alpha, \gamma) \tag{1.31}$$

where $\delta_{i,l}$ is the Kronecker delta. Then the RHS of eq.1.30can be written as

$$K_{AP} \sum_{\gamma} [V_i^\dagger(\alpha, \gamma) G^r(i\gamma, i\gamma) V_i(\gamma, \beta) \delta_{i,j}] \tag{1.32}$$

Thus, for each slab it results:

$$\widetilde{\Sigma^r_{AP}}(\alpha, \beta) = K_{AP} \sum_{\gamma} [V^\dagger(\alpha, \gamma) Gr(\gamma, \gamma) V\gamma, \beta] \tag{1.33}$$

Since

$$G^r = V \widetilde{G^r} V^\dagger \tag{1.34}$$

it results:

$$G^r(\gamma, \gamma) = \sum_{\lambda, \mu} [V(\gamma, \lambda) \widetilde{G^r}(\lambda, \mu) V^\dagger(\mu, \lambda)] \tag{1.35}$$

By replacing $G^r(\gamma,\gamma)$ from 1.35 in eq.1.33, it results:

$$\widetilde{\Sigma^r_{AP}}(\alpha,\beta) = K_{AP}\sum_{\gamma,\lambda,\mu}[V^\dagger(\alpha,\gamma)V(\gamma,\lambda)\widetilde{G^r}(\lambda,\mu)V^\dagger(\mu,\gamma)V(\gamma,\beta)] \tag{1.36}$$

Thus, by considering only diagonal terms of $\widetilde{G^r}$, eq. 1.36 is approximated to:

$$\widetilde{\Sigma^r_{AP}}(\alpha,\beta) \approx K_{AP}\sum_{\lambda}\widetilde{G^r}(\lambda,\lambda)\sum_{\gamma}[V^\dagger(\alpha,\gamma)V(\gamma,\lambda)V^\dagger(\lambda,\gamma)V(\gamma,\beta)] \tag{1.37}$$

Finally, within the approximation of $\widetilde{\Sigma^r_{AP}}$ as a blockdiagonal matrix, eq.1.37 is approximated with:

$$\widetilde{\Sigma^r_{AP}}(\alpha,\alpha) \approx K_{AP}\sum_{\lambda}\widetilde{G^r}(\lambda,\lambda)\sum_{\gamma}[V^\dagger(\alpha,\gamma)V(\gamma,\lambda)V^\dagger(\lambda,\gamma)V(\gamma,\alpha)] \tag{1.38}$$

Therefore

$$\widetilde{\Sigma^r_{AP}}(\alpha,\alpha) \approx K_{AP}\sum_{\lambda}[\widetilde{G^r}(\lambda,\lambda)I(\alpha,\lambda)] \tag{1.39}$$

where the form factor $I(\alpha,\lambda)$ is defined as

$$I(\alpha,\lambda) = \sum_{\gamma}[V^\dagger(\alpha,\gamma)V(\gamma,\lambda)V^\dagger(\lambda,\gamma)V(\gamma,\alpha)] \tag{1.40}$$

that, exploiting the following property of the matrix V,

$$V^\dagger(i,j) = V^\star(j,i) \tag{1.41}$$

can be written as

$$I(\alpha,\lambda) = \sum_{\gamma}[V^*(\gamma,\alpha)V(\gamma,\lambda)V^*(\gamma,\lambda)V(\gamma,\alpha)] \tag{1.42}$$

in which the terms are scalar. Therefore the commutative property can be applied, thus obtaining

$$I(\alpha,\lambda) = \sum_{\gamma}[|V(\gamma,\alpha)|^2|V(\gamma,\lambda)|^2] \tag{1.43}$$

For reader's ease, the explicit dependence of energy has been omitted. As pointed out in [26], care must be taken in dealing with the lead self-energies $\Sigma^r_{S/D}$ when phonon scattering is active, in order to avoid unphysical discontinuities near the injecting boundaries. For the determination of $\Sigma^r_{S/D}$, instead of adopting an iterative procedure based on an analytical approximation of G^r in the leads as in [26], in this work a fully numerical

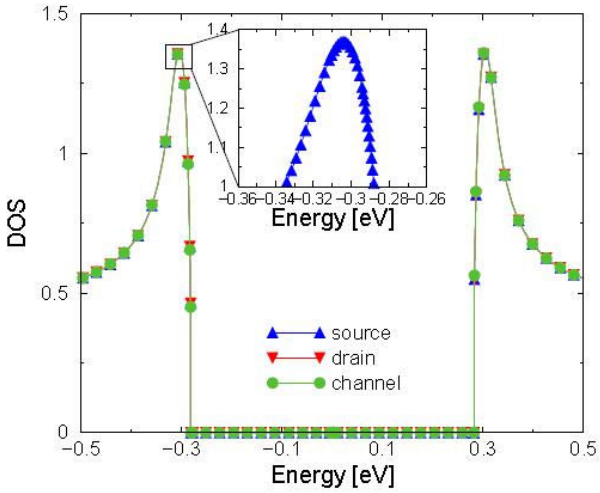

Figure 1.2: Density-of-states vs.energy computed in the presence of elastic phonon scattering in a uniform N_A = 12 intrinsic GNR in three different locations: inside the source lead, inside the drain lead, and in the middle of the device.

iteration procedure has been used. In essence, after a preliminary calculation of $\Sigma^r_{S/D}$ based on an approximation of G^r pertinent to homogeneous and infinitely long leads, the loop is entered for the self-consistent calculation of G^r in the inner domain. At each iteration step, the phonon self-energy blocks relative to the slabs adjacent to the leads are used to recalculate $\Sigma^r_{S/D}$. Then the procedure is iterated until global convergence is achieved. Fig.1.2 illustrates the LDOS vs. energy calculated in a uniform GNR at zero bias at three different longitudinal coordinates, namely within the source lead, within the drain lead, and inside the solution domain. The three overlapping curves demonstrate the consistency of the global solution. From the inset one can appreciate the effect of energy smoothing due to phonon scattering, with the suppression of the singularities at the subband edges.

As far as computational cost is concerned, in the mode-space representation the relation between phonon self-energies and Greens functions is more complex than in real space, since scattering tends to couple modes. Therefore, the possibility of a simplifying assumption to reduce the computational complexity has been tested. It consists in replacing the phonon self-energies with their respective slab-by-slab averages prior to MS conversion, thus approximating $I(\alpha, \lambda)$ in eq. 1.43 with $1/N_A$. As an example, in fig. 1.3 the turn-on characteristics of a GNR-FET with gate length equal to 17 nm and N_A= 13 are illustrated. Approximating the $I(\alpha, \lambda)$ in eq. 1.43 with a constant value leads to an overestimation of the current; instead,

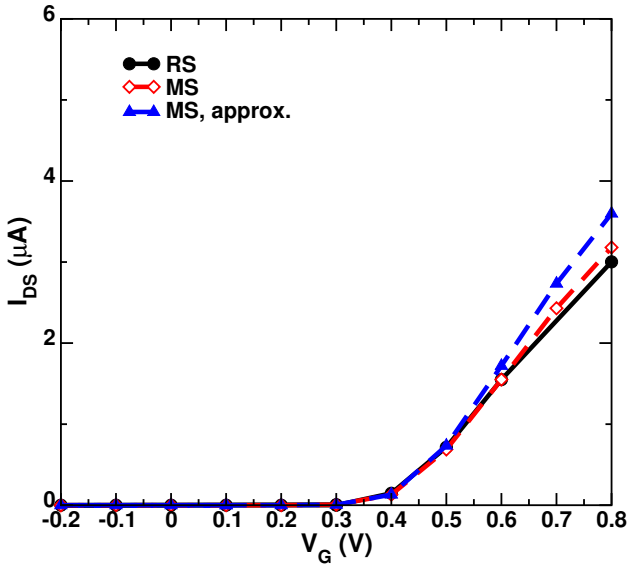

Figure 1.3: Turn-on characteristics of a GNR-FET with gate length equal to 17 nm and N_A= 13 at V_{DS} = 0.1 V, simulated accounting for AP scattering. Simulations were performed with the real space (closed circles, solid black line), mode-space with form factor fully computed (MS: open diamonds, dashed red line) or approximated to be constant (MS_approx; closed triangles, dashed blue line) approaches.

if $I(\alpha, \lambda)$ is correctly computed, the discrepancy between the mode-space and real-space approach is almost negligible. Therefore in this work, otherwise stated, the complete formulation of $I(\alpha, \lambda)$ will be used as far as the scattering with acoustic phonons is concerned.

1.2.2 Optical phonon scattering

Regarding the inclusion of optical phonon scattering, it is worth highlighting that the energy levels that differ for the energy of the optical phonon become all coupled. Therefore the calculations described above become definitely more involved, as energies can not be considered as separate, but the simulation must take into account sets af energies instead of single energy levels. Apart from the computational burden, the treatment of optical phonons is quite straightforward and an expression similar to 1.26 can be considered:

$$\Sigma^r_{OP}(i\alpha, j\beta) = K_{OP}\, G^r(i\alpha, j\beta)\delta_{ij}\delta_{\alpha\beta} \tag{1.44}$$

where K_{OP} is given by

$$K_{OP} = \frac{(D_{tk}\hbar)^2}{(2m_c\Delta E)} \tag{1.45}$$

with D_{tk} is the deformation potential, $\hbar$ the reduced Planck constant, m_c the carbon atom mass and ΔE the energy of the phonon. In this work, only optical phonons corresponding to a phonon energy ΔE equal to 160 meV [27] have been included. Fig. 1.4 depicts the turn-on and output characteristics of an armchair GNR-FET with double-gate geometry and with 2.5 nm-thick HfO_2 layers. N_A=16 and the gate length is equal to 30 nm. Source and drain regions are 30 nm long, with a doping density of 0.5 dopant per nm. Since in [33], devices with the same features were investigated by means of a real space approach, results in [33] can be taken as a benchmark to verify the accuracy of our model. From the current-voltage characteristics shown in fig. 1.4, one can assert that our simulations, performed in the mode-space approach, very well reproduce the results obtained for the full real-space even at high gate/drain voltage. The difference between the models is slightly larger at high drain voltage, where the coupling between modes becomes more involved. The phonon contribution at different bias can be understood by plotting the current flow. In figs.1.5 and 1.6 the current densities vs. longitudinal coordinate for the device presented above, at V_{DS}= 0.3 V for different gate voltages, namely V_G= 0.4 V and V_G= 0.6 V, are illustrated. As observed in [33], while at low gate voltage the contribution of optical phonons is negligible, it becomes more important at higher V_G, since a significant number of injected carriers have empty states available for scattering. Therefore an optical phonon can be emitted. Since as gate voltage increases, the number of available states increases, scattering with optical phonons is expected to affect performance more at higher gate voltages, as shown in fig. 1.4, lower panel.

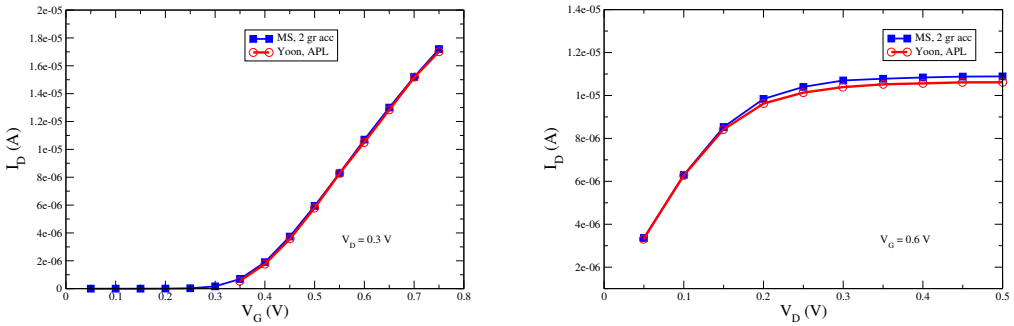

Figure 1.4: Turn-on characteristics at V_{DS} = 0.3 V and output characteristics at V_{DS} = 0.6 V of a GNR-FET with N_A= 16. The geometry of the simulated device is the same as proposed in [33], accounting for AP and OP scattering. Both our simulation results (blue line, closed symbols) and results in [33] (red line, open symbols) are reported.

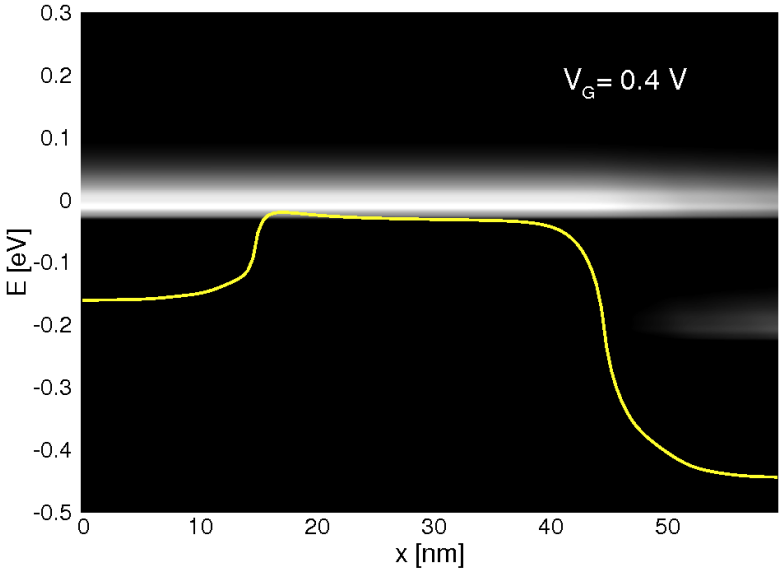

Figure 1.5: Computed current density at V_G= 0.4 V, V_{DS}= 0.3 V for a GNR-FET with N_A= 16 with the same parameters as listed in [33], with the activation of acoustic and optical phonon scattering.

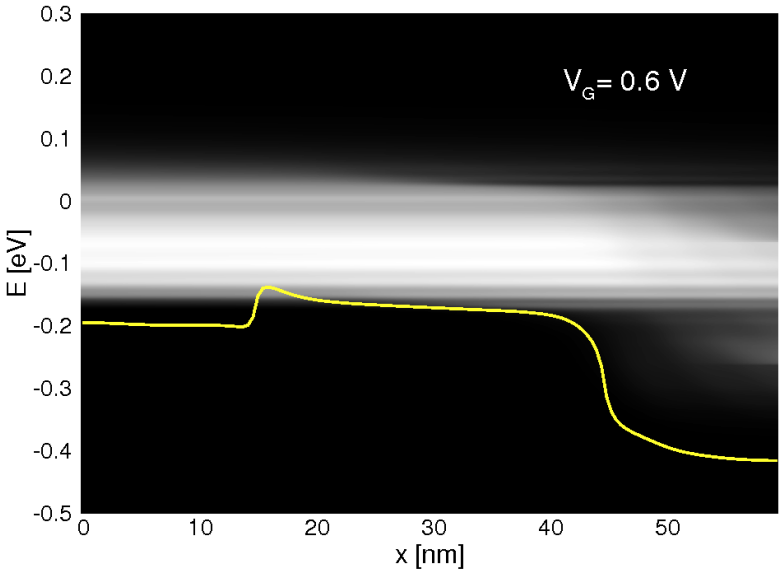

Figure 1.6: Computed current density at V_G= 0.6 V, V_{DS}= 0.3 V for the same device simulated in fig.1.5.

1.3 Self-consistency of potential solution

Finally, the electrostatic potential energy $U_{l\alpha}$ entering into the Hamiltonian is calculated by self-consistently solving the 3D Poisson equation. The box integration method is used on a discretization grid of prismatic elements with a triangular base, matching the hexagonal graphene lattice. The electron and hole charge given by 1.13 and 1.14 is directly assigned to the box surrounding the (l; α) atom. Otherwise stated, the self-consistency of the potential solution has been achieved with a precision of 0.01%.

1.4 Summary

In this chapter, the formulation of the tight-binding NEGF method has been presented, along with considerations above the inclusion of the electron-phonon interaction. Given the high computational cost of the model, reasonable approximations to improve speed-up have been discussed.

Chapter 2

GNR-FETs for digital applications

In recent years, graphene nanoribbons (GNRs) have been the object of large investigation for nanoelectronic applications, as they potentially combine the attractive features of graphene with the possibility of energy gap engineering [16] due to quantum confinement. Besides, since the energy band-gap, as previously discussed, scales inversely with the width of the GNR, only very narrow GNRs (1 to 3 nm-wide) have been proposed to be suitable for digital applications that require high I_{ON}/I_{OFF} ratio. On the other hand, the inevitable edge roughness (ER), related with the technological issue of controlling the edges with atomic precision, seriously reduces the performance of GNR-FETs [29]. Given the difficulty of experimental verification, numerical simulation can greatly help to understand the essential features of transport in GNR devices in the presence of ER as well as of other defects and scattering centers. Recently, the GNR mobility has been studied by means of semi-classical approaches [30, 31, 32] and full GNR-FETs simulations accounting for ER have been carried out using quantum transport atomistic methods [33, 34], showing a dramatic current reduction in the ON-state. In [35], the GNR-FET mobility has been investigated combining direct atomistic simulations of GNR-FETs affected by ER, as well as other types of point defects, by using semianalytical expressions for the phonon-limited mobility, showing that phonons and ER are the main causes of mobility reduction in very narrow GNR-FETs. In this chapter, the full-quantum transport model described in chapter 1 has been used to perform a simulation study on the transport properties of 1 to 2.5 nm-wide GNR-FETs. For GNR-FETs with perfectly smooth edges, the simulations have been performed in the mode-space approach (see chapter 1), which is a very good trade-off between numerical accuracy and simulation speed-up, while for GNR-FETs affected by edge roughness, the real space formulation has been adopted, to fully account for mode coupling [21]. In section 1, the device under investigation will be presented; the results of the simulations with the inclusion of elastic phonons will be discussed in

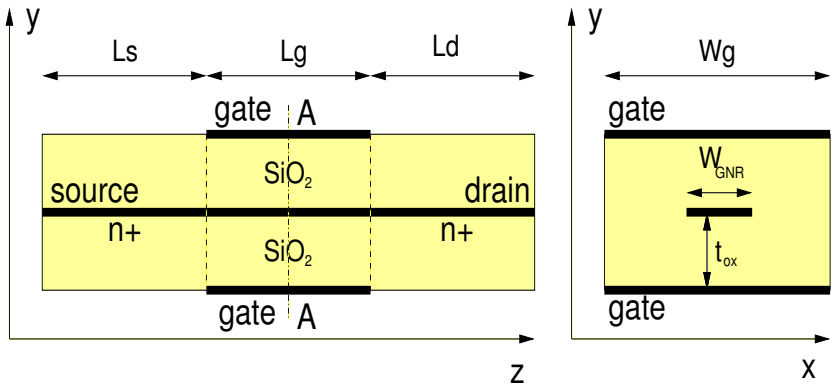

Figure 2.1: Longitudinal and transverse cross-sections of the simulated GNR-FETs.

sec. 2, while in sec. 3 the performance of GNR-FETs with rough edges will be investigated.

2.1 Simulated structure

The simulated structure is presented in fig.2.1: the device presents a double-gate topology with very thin oxide layers (silicon oxide thickness t_{ox} has been set to 1 nm), in order to maximise the electrostatic control over the channel. The symmetric 10 nm-long source and drain regions have been simulated by considering semi-infinite doped leads, while the channel has been assumed intrinsic. The doping concentration for source and drain regions has been fixed at 10^{-2} dopant atoms/carbon atoms. The nominal number of atomic dimers in the slab N_A, that defines the GNR width (W_{GNR}), and the gate length L_G have been taken as variable parameters in the study.

2.2 Effect of acoustic phonons

The inclusion of acoustic phonons (AP) in the transport model results in the suppression of the singularities of the density of states at the edge of the subband, as shown in fig.1.2. This induces a smoothing effect on the profile of the density of states, as apparent in fig. 2.2, that depicts the computed local density of states of a GNR-FET with N_A= 13 and L_G= 17 nm at a given bias with and without the inclusion of acoustic phonons. The energy-smoothing effect due to phonons is clearly visible, especially in the quasi-bound states of the valence band within the channel.

Possible effects deriving from phonon confinement due to small width of

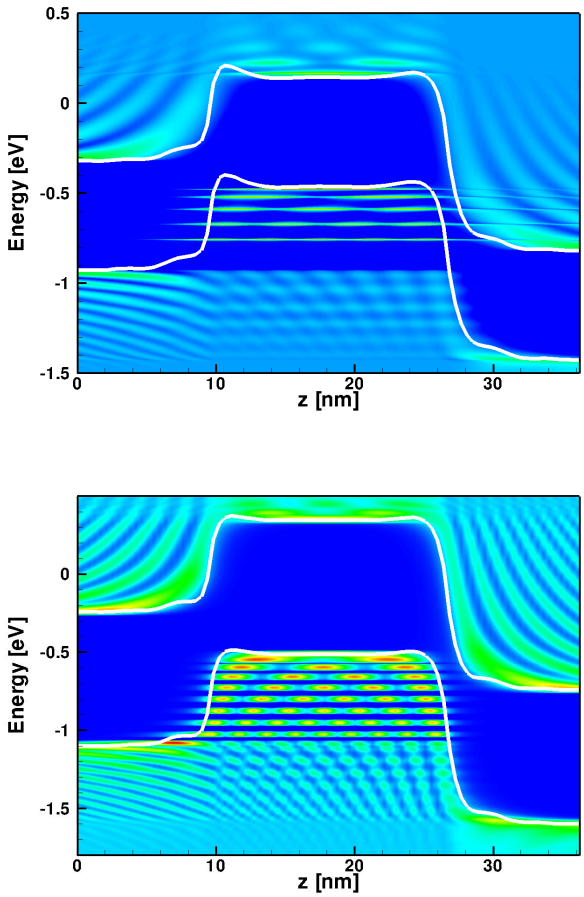

Figure 2.2: Local density of states within a GNR-FET in the ballistic regime (upper panel) and with elastic phonon scattering (lower panel) for a GNR-FET with N_A= 13 at V_G= 0 V, V_{DS}= 0.5 V.

the ribbon, and from the interfaces with the underneath and top gate insulator layers are ignored. The latter assumption is justified in case of suspended GNRs. Fig. 2.3 illustrates the turnon and output characteristics of a GNR-FET with N_A= 13 and L_G= 20 nm, computed with and without the inclusion of phonon scattering, in the case of ideal edges. From the comparison with the ballistic case, it is observed that the effect of AP scattering is not negligible, in spite of the small gate length. From fig. 2.3, the ballisticity ratio, calculated at V_G= 0.8 V, results as small as about 0.6. The reason for this can be attributed to the small width of the GNR (W= 1.5 nm), and to the phonon scattering rate, that is expected to be inversely proportional to the GNR width [30]. For completeness of the study, fig. 2.4 depicts the turnon characteristics at V_{DS}= 0.5 V and 1 V. Although AP scattering impacts significantly the current performance, the turnon characteristics at

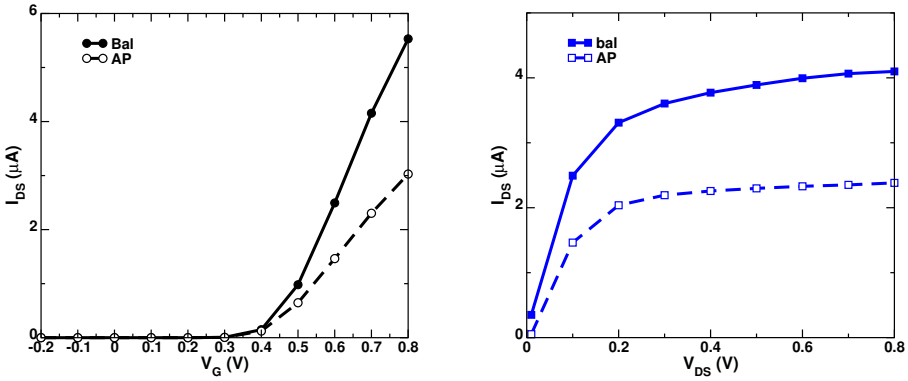

Figure 2.3: Turnon characteristics at V_{DS}= 0.1 V (left panel) and output characteristics at V_G= 0.6 V (right panel)for a GNR-FET with N_A= 13 and L_G= 20 nm simulated under ballistic conditions (solid lines) and accounting for AP scattering (dashed lines).

V_{DS}= 0.5 V shown in fig. 2.4 and the good current saturation performance with respect to V_{DS}, shown in the right panel of fig. 2.3, confirm the great potential of narrow GNR-FETs for digital applications. At very high drain voltages, remarkable accumulation of holes in the channel due to band-to-band tunneling occurs. As a consequence of this phenomenon, it is not possible to turn the device off, as evident in fig. 2.4 for V_{DS}= 1 V. For wider nanoribbons the performance limitations due to holes pile-up in the chan-

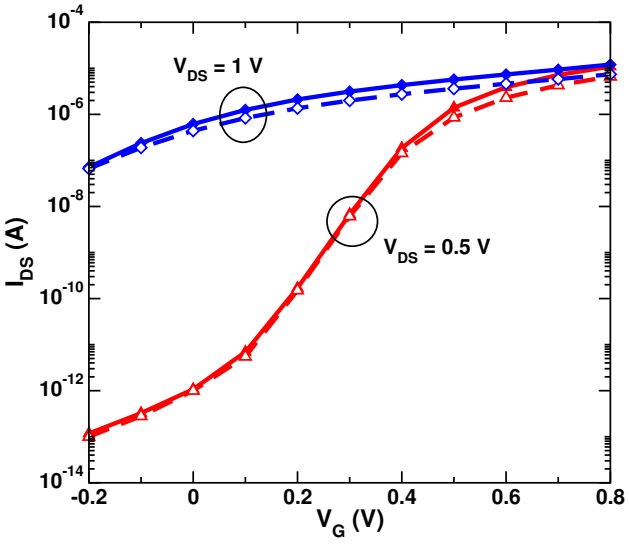

Figure 2.4: Turnon characteristics at V_{DS}= 0.5 V (red triangles) and V_{DS}= 1 V (blue diamonds) in ballistic conditions (solid lines) and accounting for AP scattering (dashed lines). The device parameters are the same used in fig. 2.3.

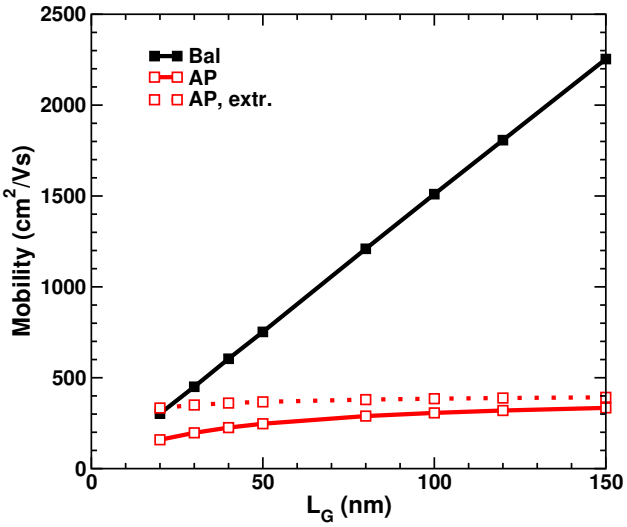

Figure 2.5: Effective (red open symbols, solid line), ballistic (black closed symbols, solid line) and AP-limited (red open symbols, dotted line) mobilities at low-field as a function of the gate length for a GNR-FET with N_A = 13 and V_{GS} = 0.6 V.

nel become more severe, as will be discussed in the next chapter.

The effective low-field mobility can be extracted from the current-voltage characteristics of devices with different gate lengths, as illustrated in fig. 2.5 for a GNR-FET with N_A= 13. The drain voltage has been set to 10 mV. This results in the device operating in linear regime. In this work mobility has been calculated as a function of gate length from the expression

$$I_{DS} = \frac{\mu Q_{ch} V_{DS}}{L_G} \tag{2.1}$$

where Q_{ch} is the average electron charge in the channel per unit length. A strong mobility reduction induced by phonon scattering is apparent. In addition, from the mobilities calculated under ballistic conditions and with AP scattering, the AP-limited mobility is extracted vs. gate length assuming the validity of Matthiessen's rule (dotted line in fig. 2.5). The extracted long-channel values are collected in fig.2.6 as a function of N_A for V_G= 0.6 V, corresponding to the device operating in ON condition. The extracted mobility values turn out to be almost proportional to N_A and divided in two sets according to the different family of seminconducting GNRs, namely the ones with N_A= 3n and 3n+1, respectively. As expected, mobility increases with device width in both families.

As far as the mobility dependence on V_G is concerned, the increase of gate voltage results in a mobility reduction, as illustrated in fig. 2.7. This can be explained on the basis of the Landauer conductance expression, which

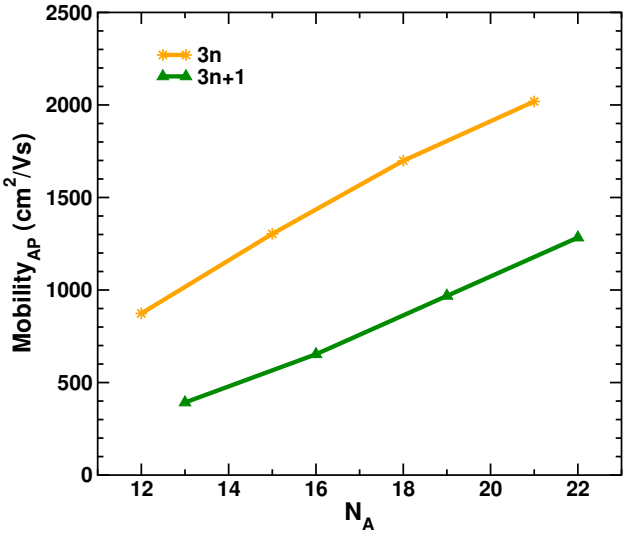

Figure 2.6: Extracted effective AP-limited mobility as a function of N_A at V_G= 0.6 V. The orange and green lines connect the points belonging to the family of N_A= 3n and 3n+1, respectively.

holds for elastic scattering processes. In fact, when the channel is driven more and more into degenerate conditions, the carrier concentration in the channel increases with V_G faster than the conductance, which ultimately tends to saturate at high gate bias. Hence, mobility decreases.

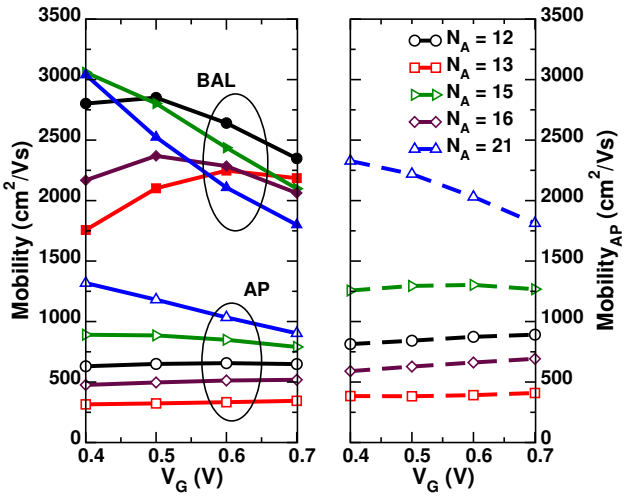

Figure 2.7: Mobility in ballistic conditions and with AP scattering (left panel) and extracted effective AP-limited mobility (right panel) as a function of V_G for GNR-FETs with different N_A.

2.3 Effect of edge roughness

As proposed in [36], edge roughness is treated in the model by adding or removing dimers independently along each edge of the GNR according to a predefined probability P, that in this study is set equal to 5%. To take partially into account the statistical variance of the results, each simulation has been repeated five times, each time with a different ER implementation. Due to the computational cost, that is quite high even for very narrow GNR-FETs, the self-consistent solution with Poisson equation has not been sought. Instead, the electrostatic potential has been determined from the self-consistent ballistic simulation of the nominal GNR-FET with perfectly smooth edges. Fig. 2.8 collects simulation results for devices with rough edges, with nominal N_A= 13 and 21, with and without the inclusion of elastic phonon scattering. The reference bias is V_G= 0.6 V, V_{DS}= 10 mV. Interestingly, for the narrower device, when ER only is active, the current decreases almost exponentially with respect to L_G, much faster than the $\propto 1/L_G$ ohmic law, which suggests the presence of a strong localization regime due to quantum interference. When AP scattering is included in the picture, the effect is to break quantum coherence, therefore the current increases. Moreover, the variance is significantly reduced. Thus, a nearly,

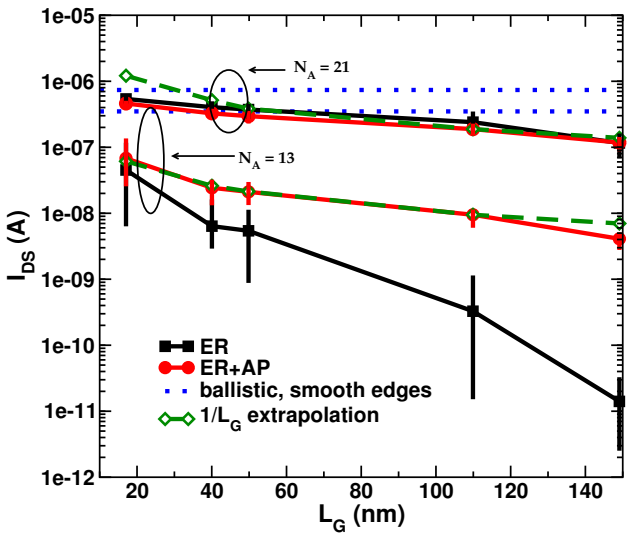

Figure 2.8: Current vs. gate length of N_A = 13 (blue diamonds) and N_A = 21 (red circles) GNR-FETs at V_G = 0.6 V, V_{DS} = 10 mV, simulated with edge roughness (ER) with P = 0.05 (solid black lines, closed squares), with ER + AP scattering (solid red curve, closed circles), and in the ballistic regime with smooth edges (dotted lines). The green dashed curves are $1/L_G$ extrapolations through the L_G = 110 nm ER+AP curves. The error bars indicate the min/max values.

but not fully, diffusive regime is recovered, as indicated by the comparison with the dashed curve, which is a $1/L_G$ extrapolation. Similar results have been found e.g. in [37], focused on the investigation of silicon nanowires with very small cross-sections. For the GNR with N_A= 21, the impact of the edge roughness is still not negligible, but the dependence on the gate length is weaker. Also, ER has much less impact with respect to the narrower case, as indicated by the quite small variance as well as by the fact that the current levels in ER + AP condition are lower than in the ER case. This is expected in a diffusive transport regime where different scattering mechanisms combine to limit current. The results for the N_A = 13 and N_A = 21 GNR-FETs at V_G = 0.6 V are $\mu \approx 17$ cm^2/Vs and $\mu \approx 307$ cm^2/Vs, respectively. When compared with the ones in fig. 2.6, these data confirm the significant impact of ER on mobility. In [29] experimental μ equal to 174 cm^2/Vs was reported for a 2.5±1 nm wide-GNR with L_G= 110 nm, which is smaller than the computed data for the N_A = 21 device of comparable width and length. A number of reasons can explain the difference, ranging from geometrical uncertainties to the presence of a variety of other scattering centers not included in the model.
For very narrow GNRs, since the current depends almost exponentially on gate length, it is not possible to define an ER-limited mobility and the use of Matthiessen's rule becomes not suitable. However, the inclusion of elastic phonon scattering recovers the diffusive law and reduces the variability. Therefore, it is possible to extract an equivalent effective ER-limited

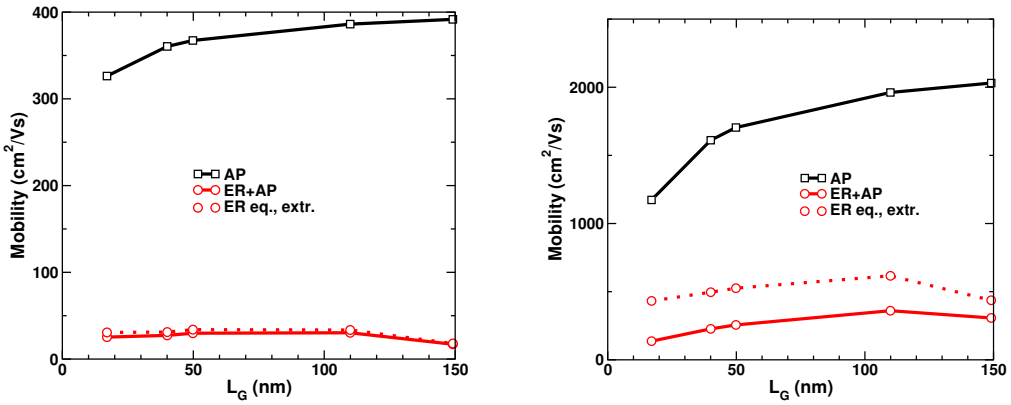

Figure 2.9: Extracted effective ER-limited mobility (dotted line) vs. gate length for GNR-FETs with different N_A=13 (left panel) and N_A=21 (right panel). Low-field mobility data for perfectly smooth (black lines, squares) and rough (red lines, circles) GNR-FETs with AP scattering are reported. The extracted effective ER-limited mobility data have been calculated by subtracting the AP data (that include ballistic component) from ER+AP data by using the Matthiessen's rule.

mobility by using the Matthiessen's rule over the mobility data in case of smooth and rough GNR-FETs with the inclusion of AP. In fig. 2.9 the extracted effective ER-limited mobility for GNR-FETs with N_A= 13 and N_A= 21 are reported. It is apparent that edge roughness affects performance definitely more than elastic phonon scattering. Finally, the study has been extended to GNR-FETs with different N_A. Because of the high computational cost of the simulations, the approach adopted in this section has been to perform simulations at selected energies, without striving for self-consistency of the solution. In particular, transmission coefficients at two energy levels have been computed. The chosen energy levels have been E= 0.1 eV (with respect to the energy reference, that is the Fermi level in the source), and the energy that corresponds to the kinetic energy equal to 0.1 eV with respect to the bottom of the conduction band in the channel. Therefore, while the first energy level is fixed for different N_A, the second changes, according to the different electrostatics of the system. The advantage of this simulation approach has been to provide a significant improvement of the statistic, moving from 5 to 50 repetitions per data point. The obtained results are illustrated in fig. 2.10. Since the variance of $\langle ln\, T\rangle$ is small, the errors bars are not reported in the picture. That also confirms that strong localization regime occurs in very narrow rough GNR-FETs even at short lengths, as it holds:

$$\frac{\Delta ln\, T}{\langle ln\, T\rangle} \ll 1 \text{ and } \frac{\Delta\, T}{\langle T\rangle} \gg 1 \tag{2.2}$$

However, as shown in the left panel of fig. 2.10, the partial recovery of diffusive law gradually occurs as width increases, for GNRs belonging to the same family. In fact, devices with N_A= 12 show strong localization, that progressively decreases as width increases to N_A= 15 and then N_A= 21. The dependence of the average logarithm of transmission on N_A for GNRs belonging to different seminconducting families, namely N_A= 3n and 3n+1, is more involved. In fact, the curve related to GNR-FETs with N_A= 12 report current levels higher than for the ones for N_A= 13. Data computed for kinetic energy= 0.1 eV are in line with these considerations, thus suggesting a very involved relation between transmission at any given energy and N_A. That results in a quite intricate dependence of the current on GNR width, as far as GNRs with rough edges are considered.

To complete the study, the localization length ξ has been calculated for the GNR-FET with N_A=13, for which the strongest localization regime at any given computed energy has been observed. In this regime, a well-defined statistical quantity is provided by $\langle ln\, T\rangle$ with

$$\langle ln\, T\rangle \approx -\frac{L}{\xi} \tag{2.3}$$

where ξ is the localized length [38]. In equation 2.3, L is the length of the disordered region; hence, in this study it is equal to the gate length. The localization length has therefore been extracted through numerical fitting for five different energies, corresponding to kinetic energy varying from 0.1 eV to 0.5 eV. The plots of the transmission coefficient with respect to gate length and the fitting curves are presented in fig. 2.11, while the fitting values are collected in table 2.1. One can observe that longer localization lengths correspond to higher energy levels, as the number of active channels increase. In table 2.2 the fitted values for ξ for different energies and N_A are summarized. While for GNR-FETs with N_A= 13 the ξ is quite low for both the computed energies, it significantly increases for N_A= 12. Confirming the considerations presented above regarding current performance for N_A= 13 and N_A= 21 devices, for both families ξ at a given energy increases as GNR width increases. In [38], focused on a simulation study on armchair GNRs with N_A= 27 and roughness probability set to 7.5% for different types of disorder, extracted localization lengths were in the range between 10 and 40 nm, according to the disorder profile, thus suggesting a very small robustness of armchair GNRs to ER. In [38], the large variability of the extracted data was also due to the fact that in the energy region where only one channel is active, the strongest deviations were observed. Moreover, as the roughness probability inversely impacts the performance, extracted mean free paths (and localization lengths) varied of more than one order of magnitude when the roughness probability

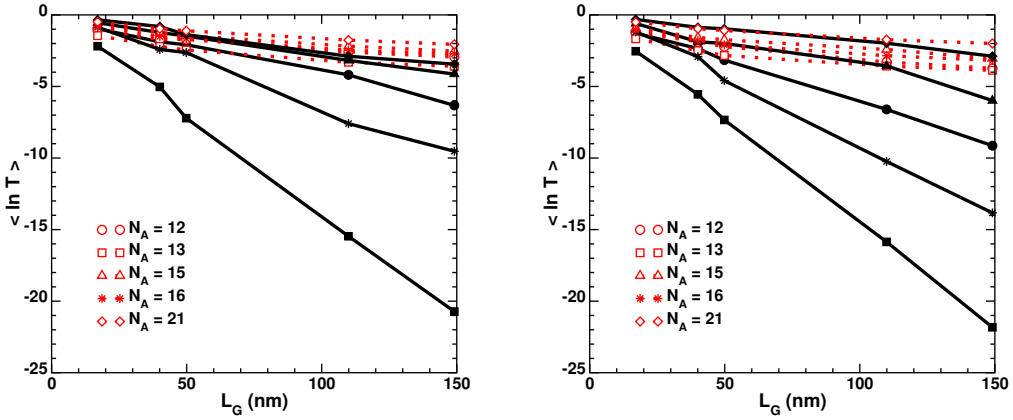

Figure 2.10: Averaged logarithm of the transmission coefficient vs. gate length for GNR-FETs with different N_A ranging from 12 to 21 at energy= 0.1 eV (left panel) and kinetic energy= 0.1 eV (right panel) at the reference bias point V_G= 0.6 V, V_{DS}= 10 mV. Solid lines with closed symbols mark ER data, while dotted lines with open symbols indicate ER+AP data. As the statistic is quite accurate, the error bars are not reported in the picture.

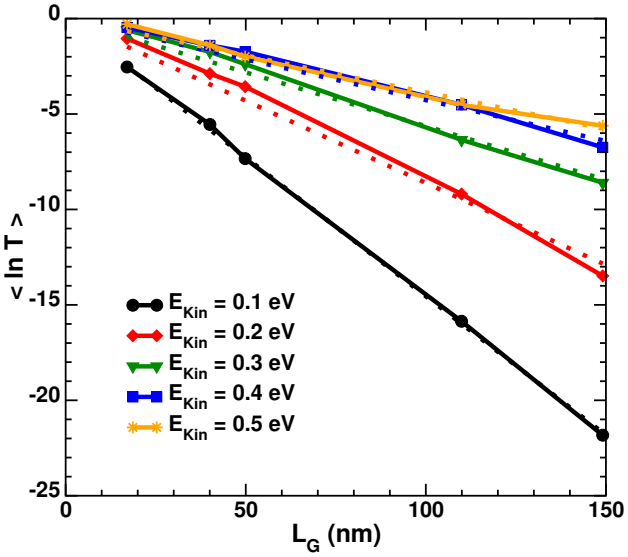

Figure 2.11: Transmission coefficient vs. gate length of N_A = 13 GNR-FETs at V_G = 0.6 V, V_{DS} = 10 mV for selected energies corresponding to kinetic energies ranging from 0.1 eV to 0.5 eV (solid curves, closed symbols). The fitting curves are shown in the same colour, by using dotted lines.

$E_{kin}[eV]$	$\xi[nm]$
0.1	6.9
0.2	11.6
0.3	17.7
0.4	23.4
0.5	25.9

Table 2.1: Fitted values for localization length ξ for GNR-FETs with N_A= 13, for different energies corresponding to kinetic energy (E_{kin}) ranging from to 0.1 to 0.5 eV.

N_A	$\xi_{E=0.1eV}[nm]$	$\xi_{E_{kin}=0.1eV}[nm]$
12	24.2	16.4
13	7.2	6.9
15	35.2	26.3
16	15.5	10.9
21	41.2	52.5

Table 2.2: Fitted values for localization length ξ for GNR-FETs with different N_A, for two chosen energy levels E= 0.1 eV and E_{kin}= 0.1 eV.

was changed from 2.5% to 7.5%.
Considering the large variability of parameters, the localization lengths extracted in this study, which are in the order of several decades of nm, are in line with data available in literature.

2.4 Summary

In this chapter an investigation on current-voltage characteristics and low-field mobility in very narrow GNR-FETs (W $\leq$ 2.5 nm) has been presented, comparing the effects of edge roughness and acoustic phonons on the performance of those devices. Full-quantum atomistic simulations have been performed, particularly oriented to the investigation of the low-field transport regime. Although the effect of elastic phonon scattering on current is not negligible even at small gate lenghts, very narrow GNR-FETs show good potential to be used in digital applications, due to their high I_{ON}/I_{OFF} ratio and good current saturating behaviour at high drain voltage. Regarding the impact of edge roughness, for very narrow GNR-FETs quantum localization effects due to ER are apparent. When AP scattering is activated, a diffusive regime is partially recovered and the statistical variance of the current is also strongly reduced. This represents good news with respect to other studies which, from simulations of GNR-FETs with ER only, had probably overestimated the current variability. However, for such narrow GNR-FETs edge roughness remains the main current limiting effect, degrading mobility much more severely than elastic phonon scattering. When phonon scattering is not activated, for very narrow rough GNR-FETs it is not possible to extract ER-limited mobility data. However, equivalent edge roughness-effective mobility values for different N_A have been provided, by using the Matthiessen's rule over rough GNRs when acoustic phonon scattering is present and a diffusive regime is recovered. In addition, localization lengths in line with data available in literature have been extracted for GNR-FETs of different widths.

Chapter 3

GNR-FETs for analog applications

It has been shown that a band-gap of several hundreds meV is necessary to achieve the on/off current ratio required in digital applications [39]. This in turn requires graphene nanoribbons (GNRs) with energy band-gaps of several hundreds of meV, shown only in very narrow GNRs (1 to 3 nm-wide) [29, 34, 30]. As discussed in the previous chapter, for such narrow nanoribbons, mobility performance is strongly limited by the edge roughness and the interaction with phonons [29, 34, 33, 35]. Moreover, successful techniques for the fabrication of those devices are far beyond the state-of-the-art technology, as the latest experiments report GNRs that are from 10 to 30 nm-wide [40, 41, 42, 43, 44]. However, these nanoribbons are less affected than narrower ones by edge roughness, and show higher mobility and better transport properties. These considerations have recently pushed the research on graphene and GNRs towards analog high-frequency applications [45, 46, 47, 9, 48, 49, 50, 51], for which achieving high cut-off frequencies is the key point and it is not required to turn the current off. In [52], an extensive study on aspects like stability, gain, power dissipation and load impedance of ballistic graphene GNR-LNAs led to the conclusion that GNRs with the energy band-gap of the order of 100 meV are ideally suited for that application. In this chapter the full-quantum trasport model described in cap. 1 has been used to perform a simulation study on 10 to 15 nm-wide GNR-FETs (which show a 90 to 140 meV band-gap), and to provide design suggestions and optimization criteria. In section 1 the simulated device and its current-voltage characteristics will be presented; sec. 2 will be focused on discussing the main causes of the non-saturating behaviour typical of these devices and design criteria will be suggested; in sec. 3 the investigation will be extended to the the simulation of GNR-FETs with metal contacts, and considerations above methods to optimize the performance will be proposed. Finally, in sec. 4 a generalization of the study will be presented, to explore the maximum performance achievable.

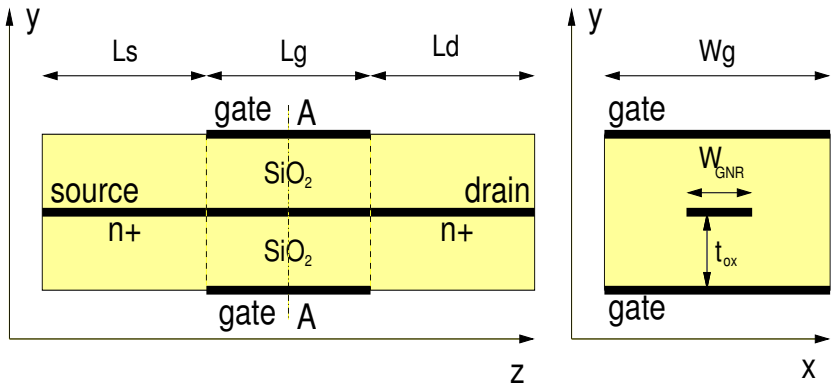

Figure 3.1: Longitudinal and transverse cross-sections of the simulated GNR-FETs.

3.1 Performance evaluation

The reference architecture for the device, previously illustrated in 2.1, is reported for convenience in fig.3.1. In this study, 10 nm-wide semiconducting GNRs with armchair ideal edges are taken as channel material, and a double-gated structure with silicon oxide 1 nm-thick is investigated. The 10 nm-long source and drain regions are doped with the same doping species while channel is intrinsic, leading to an n-i-n topology. Source and drain doping concentrations are somewhat arbitrary at this stage and will be optimized later. The channel length is short (L_G= 10 nm) to seek high-frequency performance. From the turn-on characteristics depicted in fig. 3.2, it is apparent that the device is not suitable for logic operation, since the largest on/off current ratio is nearly 5. This confirms that 10 nm-wide GNR-FETs are not suitable for logic applications.
The output characteristics presented in fig. 3.3 show the complete absence of a saturating behaviour, i.e. the output conductance

$$g_d = \frac{\partial I_{DS}}{\partial V_{DS}} \tag{3.1}$$

always remains high. The lack of a saturation region results in a strong limitation on the maximum voltage-gain achievable, given by

$$A_v = \frac{g_m}{g_{ds}}, \quad \text{where} \quad g_m = \frac{\partial I_{DS}}{\partial V_{GS}} \tag{3.2}$$

is the transconductance. Besides that, the choice of the bias point becomes critical, since the low g_d region is very narrow. The reason for the absence of a clear saturation region is illustrated in fig. 3.4 , reporting band diagrams and current energy spectra relative to the bias points marked A and B in fig. 3.3. Due to the small band-gap (E_G= 0.14 eV), at high V_{DS}

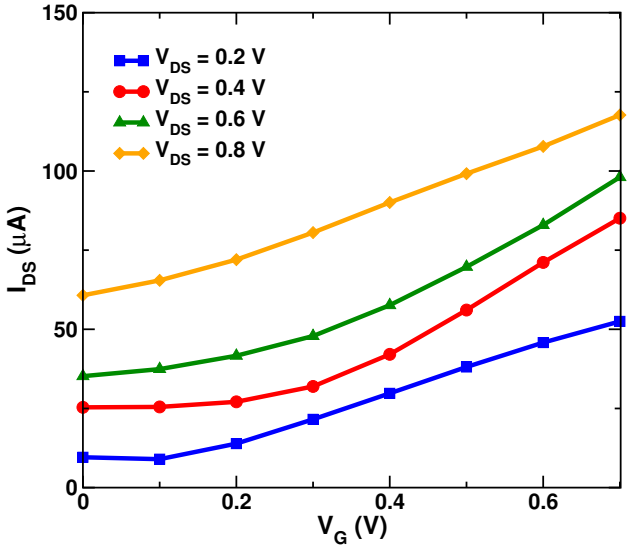

Figure 3.2: Turn-on characteristics of a simulated ideal GNR-FET with source/gate/drain lengths equal to 10 nm each, GNR width equal to 10 nm (corresponding to N_A = 82 dimers in the GNR 1D unit cell), silicon oxide 1 nm-thick, source and drain doping fractional concentrations equal to $5 \cdot 10^{-3}$.

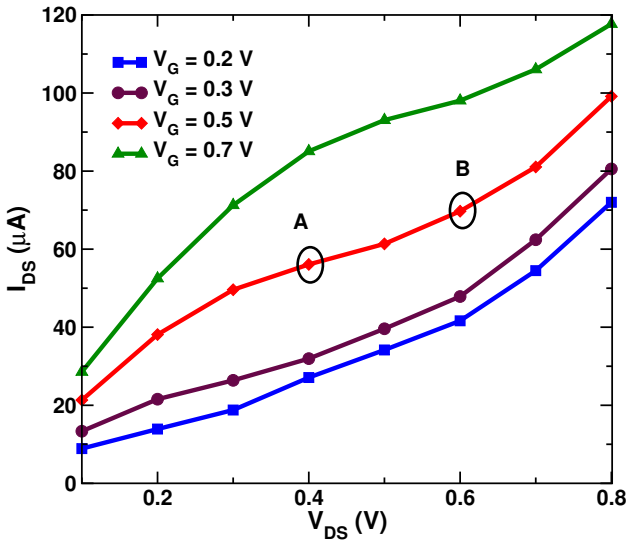

Figure 3.3: Output characteristics of the simulated ideal GNR-FET with the same parameters as in fig. 3.2. Labels A (at V_G= 0.5 V, V_{DS}= 0.4 V) and B (at V_G= 0.5 V, V_{DS}= 0.6 V) indicate the bias points relative to the plots in fig. 3.4.

the bottom of the conduction band and the top of the valence band are present in the same energy interval. Thus, band-to-band-tunneling (BTBT)

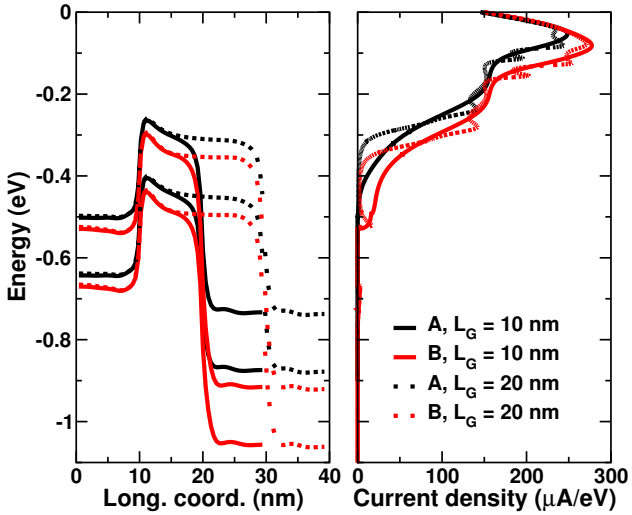

Figure 3.4: Lowest subband diagram (left) and total current spectral density vs. energy (right) for the device with gate length equal to 10 nm simulated in figs. 3.2 and 3.3 for the two bias points indicated with the labels A and B in fig. 3.3. The curves relative to the simulation of a GNR-FET with gate length equal to 20 nm are also reported. The zero energy reference is the Fermi level in the source.

at the drain-end of the channel takes place such that electrons in the valence band in the channel are almost in equilibrium with the Fermi level in the drain. Hence, increasing V_{DS} causes a depletion of electrons in the valence band in the channel, with a consequent positive charge accumulation due to the electrons in the channel escaping to the drain. That, due to the static feedback, induces the reduction of the potential energy barrier which is responsible for the current increase and substantial non-saturating behaviour. The effect is similar to a Drain-Induced Barrier-Lowering (DIBL) in standard MOSFETs, but in this case the reason is not the occurrence of short-channel effects. In fact, when the gate lenght is doubled from 10 nm to 20 nm, it is observed that the energy barrier height does not depend on L_G (fig. 3.4, left panel) and only minor variation occurs in the current density spectra (fig. 3.4, right panel). Since the lack of saturation is not due to short-channel effects (SCEs), increasing the gate length in GNR-FETs does not improve performance, unlike reported for standard silicon MOSFETs.

3.2 Design considerations

3.2.1 Asymmetrical doping

From the discussion above, one can deduce the bias conditions for which voltage-gain (expressed in eq. 3.2) and cut-off frequency, given by

$$f_T = \frac{1}{2\pi} \frac{g_m}{C_{gs}} \tag{3.3}$$

where C_{gs} is the gate capacitance, are maximised. The bias point is therefore chosen as follows: in order to have large g_m, V_G must be high enough to drive the channel into degeneracy, while, in order to keep g_d small, V_{DS} must be set at the highest possible value compatible with the absence of BTBT between the valence band in the channel and the conduction band in the drain. For the same reason, the doping level in the drain should be kept low, such that the energy difference between the Fermi level and the conduction band edge in the drain is minimised, leading to devices with asymmetrical doping in source and drain regions. In fig. 3.5, the energy bands and current spectra for the 10 nm-wide, 10 nm-long device presented in fig.3.1 are shown. The chosen bias point is V_G = V_{DS}= 0.27 V, the doping in the source region is set to N_S= 10^{-3} dopant atoms/carbon atoms and the doping in the drain region is set at the lowest value that allows the

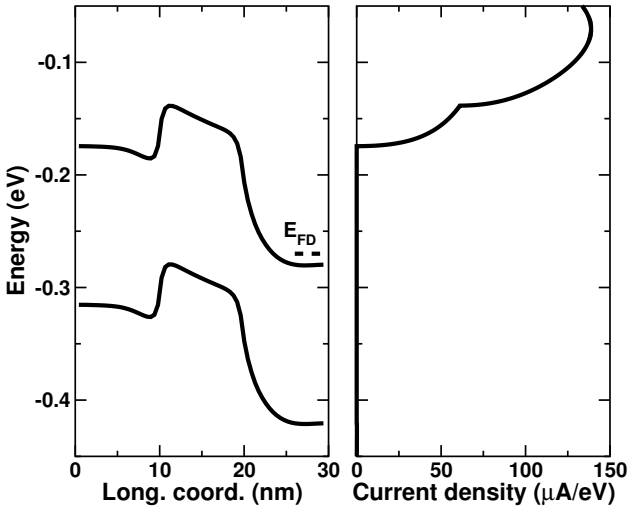

Figure 3.5: Lowest subband diagram (left) and total current energy spectrum (right) for the GNR-FET of figs. 3.2 and 3.3. The parameters are the same, except source and drain doping fractional concentrations, set at $1 \cdot 10^{-3}$ and $0.6 \cdot 10^{-3}$ respectively. The bias condition (V_G = V_{DS} = 0.27 V) is the one for which the voltage gain is largest. The dashed line indicates the Fermi level in the drain which is equal to -qV_{DS}.

correct operation of the drain, namely $N_S = 6 \cdot 10^{-4}$ dopant atoms/carbon atoms. Once the optimum doping conditions have been identified, the voltage-gain as a function of bias around the most favourable point can be calculated from the current characteristics. The results are presented in fig. 3.6, where g_m and g_d are calculated through numerical differentiation of the current with steps of 3 mV. The step is chosen as the smallest that allows numerical accuracy of the results, without them being affected by quantum oscillations. The average gain value is around 15, with a mild increase with the gate voltage, as expected from the channel being in degenerate conditions. Outside the range of V_{DS} considered in the figure (about 40 mV), the gain decreases rapidly, confirming a quite modest extension of the high-gain quasi-saturation region. Although voltage-gain performance is not high, it is sufficient for the design requirements of several analog blocks. For example, the maximum voltage gain achievable for a two-stage opamp could be around 200. The unit short-circuit current-gain frequency f_T in the same bias region is reported in fig. 3.7, where the gate capacitance is calculated in a quasi-static approximation by numerical differentiation with 3 mV steps. It is remarkable that throughout the bias window f_T is rather constant and largely exceeds the THz barrier, with a maximum value of around 7 THz, thus confirming the excellent high-frequency potential of GNRs. The reported value for voltage-gain and short-circuit current-gain frequency should be considered as theoret-

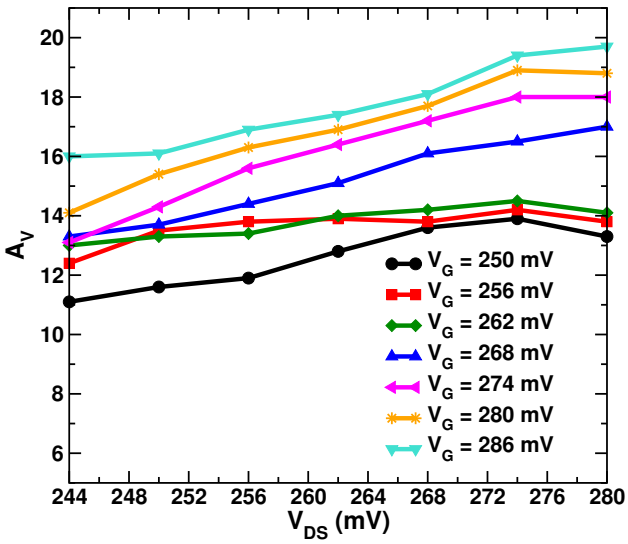

Figure 3.6: Small-signal voltage-gain vs. bias point for the GNR-FET of figs. 3.2 and 3.3 with the optimized doping concentrations presented in fig. 3.5 within a bias range of about 40 mV. The gain has been obtained calculating the transconductance g_m and output conductance g_d using finite differences with voltage steps of 3 mV.

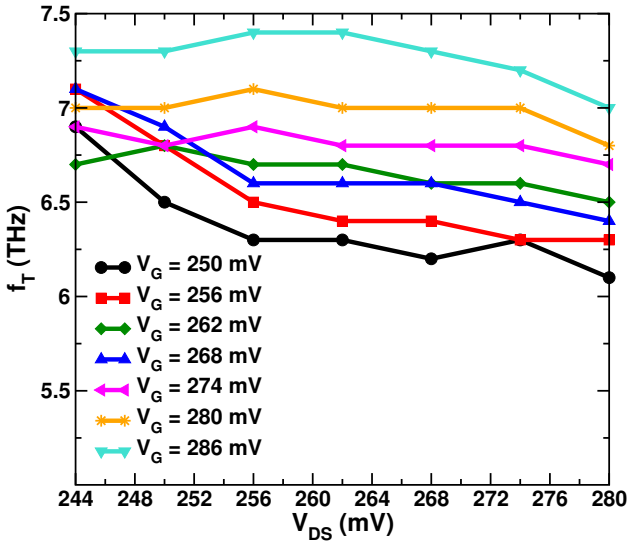

Figure 3.7: Unit short-circuit current-gain frequency vs. bias point for the same GNR-FET as in fig. 3.6 in the same bias range.

ical upper limits, since ideal ballistic transport has been assumed so far with perfectly defined GNR edges. Since the main limitation of performance in the 10 nm-wide GNR-FETs is the band-to-band tunneling, it is interesting to repeat the study for wider GNR devices, since the energy-gap scales almost inversely with the GNR width [30]. The simulations were then performed for the same architecture shown in fig. 3.1, considering GNR with a width of 15 nm instead of 10 nm. The energy-gap for these devices is only 90 meV, thus the designer should expect even more

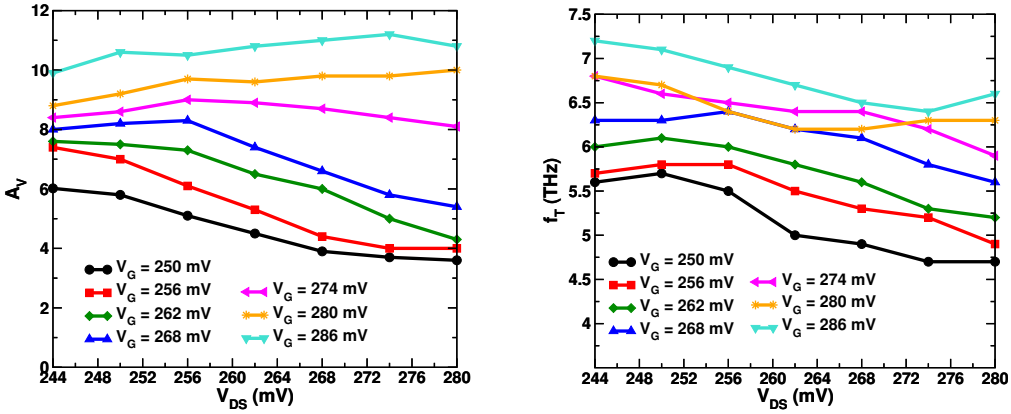

Figure 3.8: Small-signal voltage-gain (left panel) and unit short-circuit current-gain frequency (right panel) vs. bias point for the same GNR-FET as in fig. 3.6 except for the larger width of the GNR (W_{GNR} =15 nm, corresponding to N_A = 124 dimers in the GNR unit cell).

$W_{GNR}\,[nm]$	$L_G\,[nm]$	$Model$	$g_m\,[\mu S]$	$g_d\,[\mu S]$	$C_{gs}\,[aF]$	A_v	$f_T\,[THz]$
10	10	bal	116	6.6	2.7	17.6	6.8
		AP	83	6.3	2.7	13.2	5.0
	20	bal	129	6	4.6	21.5	4.4
		AP	90.3	10.3	4.8	8.7	3
15	10	bal	168.3	25.8	4.4	6.5	6
		AP	144.3	25.7	4.6	5.6	5
	20	bal	175.3	31.6	9.3	5.5	3
		AP	143	32.7	8.5	4.4	2.7

Table 3.1: Summary of the main results obtained for GNR-FETs of different L_G, namely 10 and 20 nm and different widths: 10 nm (corresponding to N_A = 82) and 15 nm (corresponding to N_A = 124). The other device parameters and bias are as in fig. 3.5 (optimized drain doping). The results both in ballistic conditions (bal) and with acoustic phonon scattering (AP) are shown. The scattering with elastic phonons is treated as in chapter 1.

severe limitations. In fig. 3.8 A_V and f_T performances are illustrated. The voltage-gain is between 5 and 10 and it shows evident degradation with respect to fig. 3.6, even though the variation in the observed region is quite small. On the contrary, f_T is very high (between 6 and 7 THz) and almost unchanged. In tab. 3.1, at the optimized bias point previously discussed (V_G=V_{DS}= 0.27 V and asymmetrical doping in source and drain regions), A_V and f_T are reported for different scenarios. The GNR width and gate length vary from 10 to 15 nm and from 10 to 20 nm respectively, with the simulations performed either in ballistic conditions or including acoustic phonon (AP) scattering within the full-quantum approach described in chapter 1. The transition frequency is always very high and presents only minor variation, while the voltage-gain varies from 4.4 to 21.5, confirming that it is rather limited and also very affected by the variability of the parameters. The main effect of the acoustic phonons inclusion is the slight reduction of g_m, which impacts both A_V and f_T.

3.2.2 Use of high-k oxide

Another viable approach to contrast the reduction of potential energy barrier given by the charge-accumulation in the channel region due to BTBT is the use of a high-k oxide to achieve better electrostatic control. In recent years high-k dielectrics have been introduced to facilitate the EOT (Equivalent Oxide Thickness) scaling and reduce leakage current. Therefore in the structure under investigation presented in fig. 3.1 with L_G = 10 nm, silicon oxide (ϵ_r = 3.9) has been replaced with aluminum oxide (ϵ_r = 9).

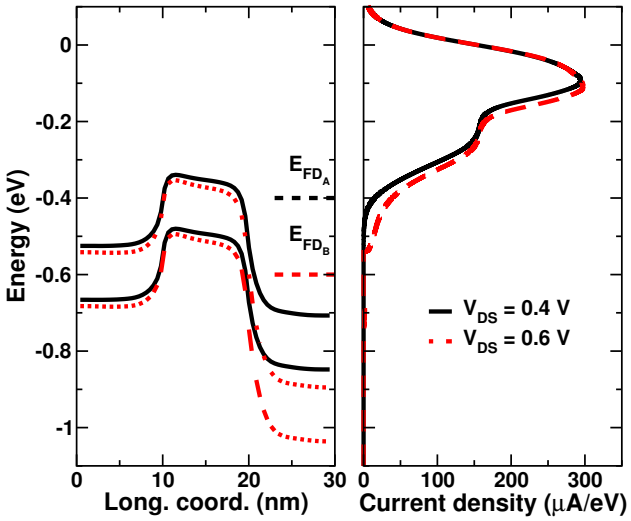

Figure 3.9: Lowest subband diagram (left) and total current spectral density vs. energy (right) for a GNR-FET with the same parameters listed in fig. 3.2, but using Al_2O_3 instead of SiO_2 as dielectric. The bias is as in fig. 3.4, which this figure should be compared with (gate length equal to 10 nm). Dashed lines indicate the Fermi level in the drain which is equal to -qV_{DS}, for drain voltage equal to 0.4 V (black line) and 0.6 V (red line).

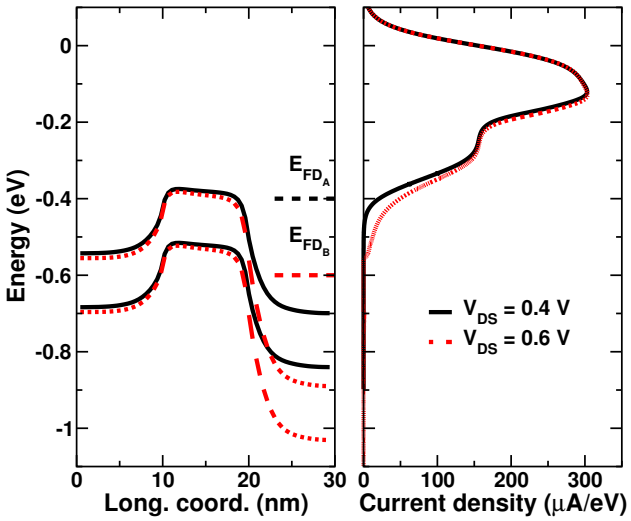

Figure 3.10: Lowest subband diagram (left) and total current spectral density vs. energy (right) for a GNR-FET with the same parameters listed in fig. 3.2, but using HfO_2 instead of SiO_2 as dielectric. The bias is as in fig. 3.4 and 3.9. As in fig. 3.9, the dashed lines indicate the Fermi level in the drain, for V_{DS} equal to 0.4 V (black line) and 0.6 V (red line).

In fig. 3.9, lowest subband diagram (left panel) and total current spectral density (right panel) of this device are reported and should be compared with fig. 3.4 as reference. The reduction of the potential energy barrier in this case is very limited and the energy profiles present weak dependence on V_{DS}. As expected, substituting aluminum oxide with hafnium oxide (ϵ_r = 16) provides additional benefit (fig. 3.10). The channel potential is even less sensitive to the drain voltage and the energy profiles show only minor dependence on V_{DS}. This results in a gain increase (A_V = 44), with a mild speed penalty (f_T = 4.7 THz) due to the higher gate capacitance.

3.2.3 Introduction of underlap region

Another possible solution to limit band-to-band tunneling is modifying the architecture of the device, by introducing an intrinsic gate-drain underlap layer that separates channel and drain regions, as shown in fig. 3.11. With increasing gate voltage a residual potential barrier in the underlap region is observable, that limits the maximum current achievable. In consideration of the degradation of the gate control over the channel potential, that ultimately leads to deteriorating the transconductance and therefore both A_V and f_T, the conclusion is that this configuration does not provide improvement with respect to the standard case.

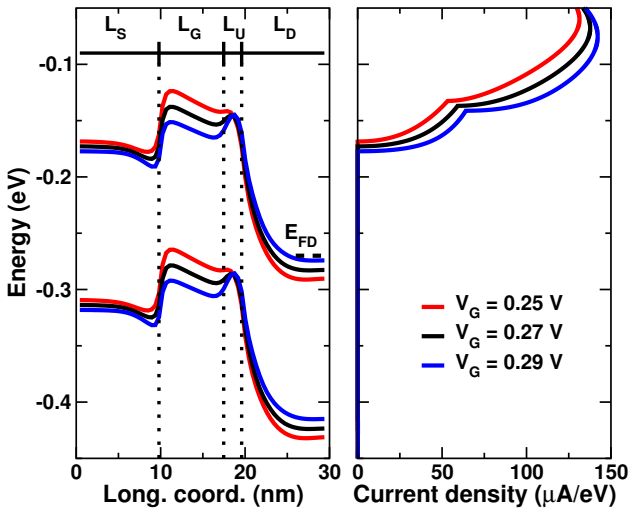

Figure 3.11: Lowest subband diagram (left) and total current spectral density vs. energy (right) for three different gate voltages: V_G= 0.25 (red line), V_G= 0.27 (black line) and V_G= 0.29 (blue line). The GNR-FET parameters are as in fig. 3.5, but the gate length has been reduced to 8 nm and an intrinsic gate-drain underlap region with length L_U = 2 nm has been introduced.

3.3 Extension to GNR-FETs with metal contacts

In order to extend the study to the analysis of more realistic devices, the simulated structure is modified as follows: the silicon oxide thickness is doubled (from 1 to 2 nm) and the gate length is increased from 10 to 40 nm. This dimension is sufficient to neglect direct source-to-drain tunneling, so to focus the analysis on the BTBT at the drain. The gate length could be reduced in a design seeking the ultimate speed performance. The source and drain region extension is 5 nm and 50 nm respectively. Besides that, the study is not limited to the assumption of semi-infinite leads, but even metal Schottky barrier (SB) contacts are considered. To reduce computational effort without loss of accuracy, SB contacts are treated as described in chapter 1, with a phenomenological approach. Namely, the electron injection from the metal into the GNR is modelled by means of a diagonal retarded self-energy for the atoms located at the source and drain contact edges. The modulus of these elements has been defined equal to 5 eV. The SB height is assumed the same for electrons and holes and equal to half the band-gap. Therefore, it is GNR-width dependent. The output characteristics of the 10-nm wide GNR-FET with metal SB contacts are illustrated in fig. 3.12. The doping in the drain region is fixed at $N_D = 5 \cdot 10^{-4}$, that is half the concentration set in the source region, so to avoid excessive

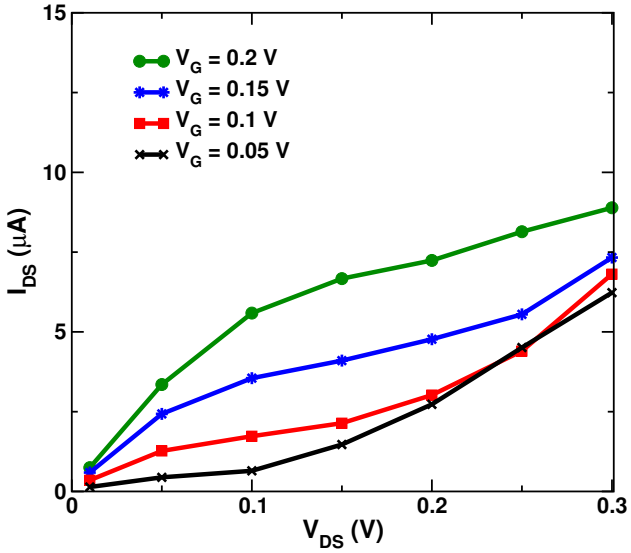

Figure 3.12: Output characteristics of the simulated GNR-FET with metal contacts, silicon oxide 2 nm-thick, source/gate/drain lengths = 5/40/50 nm respectively illustrated in fig. 3.1, except for the use of SB contacts instead of semi-infinite leads. The fractional source concentration is costant and equal to $1 \cdot 10^{-3}$, while in the drain it is equal to $0.5 \cdot 10^{-3}$.

electron degeneracy that would favor BTBT. It is observed that there is no saturation region for drain voltages higher than the energy band-gap. In GNR-FETs with semi-infinite doped leads the cause of the non-saturating behaviour is the positive charge accumulation in the channel and the induced reduction of the potential energy barrier. To investigate if the same phenomenon occurs in devices with metal contacts, the lowest subbands profile (left panel) and total current spectra (right panel) are reported in fig. 3.13. The contact injects at all energies and forms a SB with a GNR. The presence of the SBs is evident from the band-bending close to the contacts. In this scenario, the holes which are injected into the channel from the drain through band-to-band tunneling can find their way out to the source. This process is highly favoured by the very small length of the source region (5 nm) that can be easily tunneled through. As a consequence, the hole pile-up in the channel is strongly reduced and the potential barrier and the electron current are only minimally affected by the drain voltage change. On the other hand, a direct current path is present from drain to source for relatively low-energy holes assisted by a double tunneling process, that is BTBT at the drain-channel junction and intraband tunneling in the source. This leads to an overall non-negligible hole current, as observable in the right panel of fig. 3.13, dashed line. Thus, for these devices, the phenomenon that prevents current saturation is not the pile-up but the direct source-to-drain tunnel. In this scenario, careful

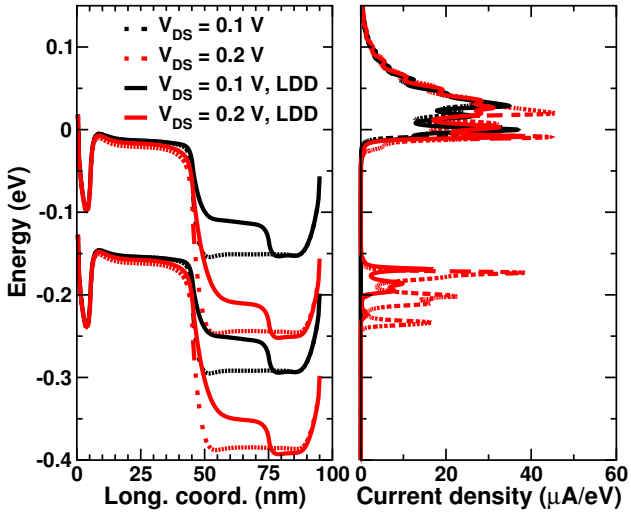

Figure 3.13: Band diagrams (left) and current spectra (right) calculated for the two bias points V_G = 0.1 V, V_{DS} = 0.1 (black lines) and 0.2 V (red lines) for the device of fig. 3.1 with (solid lines) or without (dashed lines) the introduction of a 30 nm-long LLD region. The fractional doping in the LDD extension is exponential ed equal to 2 to $3 \cdot 10^{-4}$.

band engineering increasing distance between the bands is a promising technique to limit tunneling and then improve performance significantly.

3.3.1 Optimization criteria

The effect of the introduction of a 30 nm-long light doped drain extension (LDD) is illustrated in fig. 3.13, solid lines. It is evident that the hole current is strongly modulated by the presence of the LDD, unlike reported for semi-infinite doped leads standard case in subsection 3.2.3. In fact, for the FETs with metal contacts, holes in the valence band in the channel are not necessarily in equilibrium with the Fermi level in the drain. The ideal condition is reached when the LDD provides a sufficiently long BTBT distance so as to limit the hole current as much as possible. On the other hand, the source should be nearly transparent to holes, such that holes entering from the drain do not accumulate in the channel. This is the reason for the choice of a 5-nm source adopted here. The complete elimination of the source region, with the metal contact aligned to the gate edge, would lead to the so called SB-GNRFET, where the gate voltage modulates the tunneling distance through the SB seen by the carriers injected from the source. The solution adopted here with doped source is preferable in terms of control over the current flow. With regard to general guidelines for the LDD engineering, the first consideration is that increasing the length of the LDD region would limit transit time and then cut-off frequency. This imposes a trade-off between LDD length and speed. The choice of a LDD= 30 nm is here dictated by the limit of having the LDD length comparable with the gate length. Secondly, the best criterion to choose the doping profile in the LDD region is to fix it at the value that results in a potential profile as close to linear as possible within the LDD itself for a given gate and drain voltage. That maximises the tunneling distance for all energies between the bands, therefore is the one that leads to the best performance. Clearly, that is impossible to satisfy over the full bias range, since the potential profile in the LDD is determined by the doping as well as by the electron concentration, and the latter is bias dependent. So a preliminary choice should be made on which portion of the I-V characteristic should be privileged. In fig. 3.14 an example is given. The bias point is V_G = V_{DS} = 0.2 V, chosen since it is the one for which g_m is highest. In order to have an approximately linear behaviour of the electric potential within the LDD, it is necessary to define a donor doping profile that matches as close as possible the electron concentration, so as to give zero space charge as indicated in figure 3.14 (dashed line). In general, an exponential doping shape is an acceptable compromise between effectiveness and simplicity. In the case illustrated in the figure, the exponential profile is weak (N_{LDD}

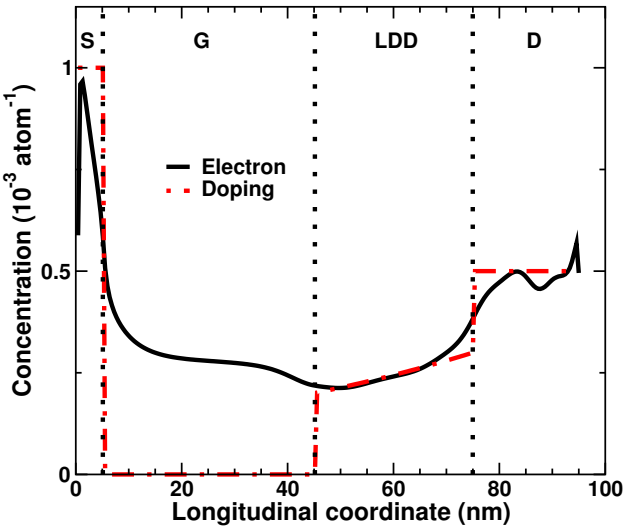

Figure 3.14: Fractional electron concentration (black solid line) calculated at V_G = V_{DS} = 0.2 V for the GNR-FET of fig. 3.13 with SB contacts and LDD extension. The optimized doping profile is also shown (red dashed line).

varies from 2 to $3 \cdot 10^{-4}$ dopant atoms/carbon atoms), then it looks almost linear. Fig. 3.15 shows the output characteristics of the GNR-FET with the modified doping concentration and should be compared with fig. 3.12, that was obtained without the introduction of the LDD region. The output characteristics show wider operation region with respect to the ones

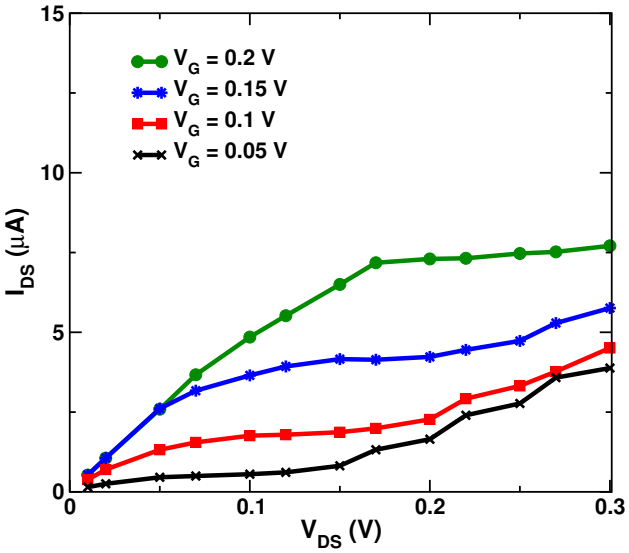

Figure 3.15: Output characteristics of the simulated GNR-FET with SB contacts, LDD extension and the optimized doping profile shown in fig. 3.14.

obtained without LDD and evidently better saturation performance. As expected, the improvement is not obtained for all bias conditions and is more effective in the bias range 0.15 V $<$ V_G $<$ 0.2 V, V_{DS}= 0.2 V, that is close to the bias for which the doping profile was optimized. The explanation of the reason why at low drain voltages the characteristics at high V_G overlap is given in fig. 3.16, that depicts the energy band diagrams at V_G from 0.05 to 0.2 V, V_{DS}= 0.05 V (left panel) and 0.2 V(right panel). The green line in the right panel corresponds to the optimized band profile, that is almost linear as expected. For that doping that maximises the linearity of band profiles at V_G= V_{DS} = 0.2 V, at lower drain voltages (left panel) for V_G= 0.15 V the creation of a potential energy barrier in the LDD region strongly limits the gate control over the current. This explains why in fig. 3.15 the current curves for V_G= 0.15 V and V_G= 0.2 V at low drain voltage almost overlap. Moreover, in the right panel it is noticeable that at gate voltages lower than 0.2 V, the band profiles rapidly depart from the linear shape, which is detrimental for BTBT and explains why the optimization performed at the selected bias is not very effective at low V_G. The small-signal device transconductance g_m at V_{DS} = 0.15 V is shown in fig. 3.17. The small wiggles in g_m are due to quantum mechanical resonances, while the negative value at low VG is due to the typical ambipolar behavior. The peak value normalized to the GNR width is almost 6 mS/μm, quite high value which confirms the excellent transport properties of graphene although it should be reminded that phonon and defect

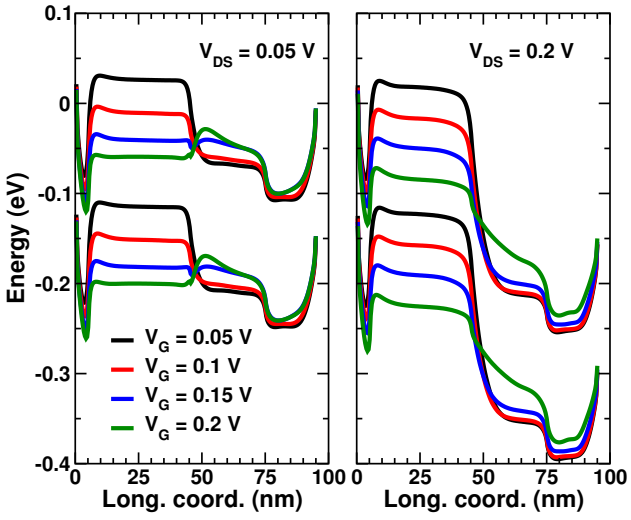

Figure 3.16: Band diagrams for the GNR-FET with SB contacts, LDD extension and optimized doping profile for V_{DS} = 0.05 V (left) and V_{DS}= 0.2 V (right) for increasing V_G from 0.05V (black curve) to 0.2 V (green curve).

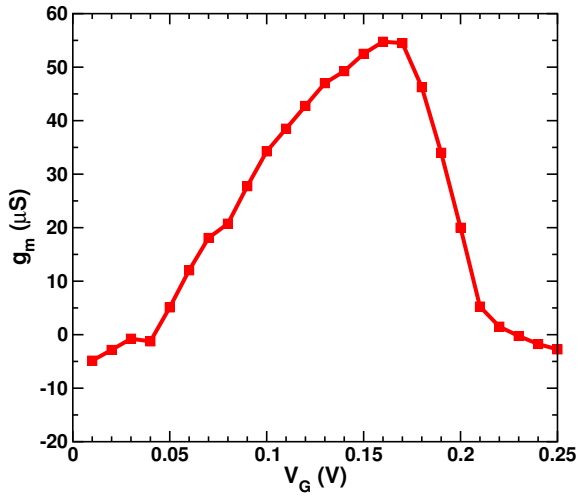

Figure 3.17: Transconductance vs. gate voltage at V_{DS} = 0.15 V of the GNR-FET with SB contacts and 30 nm-long LDD extension with optimized doping profile. The derivative is calculated by numerical approximation with steps of 20 mV.

$V_G, V_D[V]$	$g_m[\mu S]$	$g_d[\mu S]$	$C_{gs}[aF]$	A_v	$f_T[THz]$
0.05,0.07	13.9	1.9	3.7	7.3	0.6
0.1,0.12	34.5	2.1	5.1	16.4	1.1
0.15,0.17	53.3	1.6	6.0	33.3	1.4
0.2,0.22	64.5	3.5	6.3	18.4	1.6

Table 3.2: Small-signal performance parameters of the 10 nm-wide GNR-FET with optimized doping profile illustrated in fig. 3.14.

scattering have not been included in this study. The other small-signal parameters are collected in table 3.2, where the output conductance g_D, the gate capacitance C_{GS}, the voltage gain A_V and the cut-off frequency f_T are reported for different bias points corresponding to the best saturating portion of each curve of fig. 3.15. The voltage-gain is quite good, especially around the optimized bias point, and cut-off frequency is always very high, around 1 THz. For completeness and ease of comparison, the optimization was performed at different bias, leading to different doping profiles in the LDD region. The result is illustrated in fig. 3.18. It is evident that careful choice of parameters could lead to saturating current characteristic at the select V_G.
Finally, the study was repeated using semi-infinite leads instead of metal contacts as boundary conditions. The band diagrams and current energy spectra are reported in fig. 3.19. As discussed in subsection 3.2.3, the introduction of a lightly-doped underlap region in this case has only minor

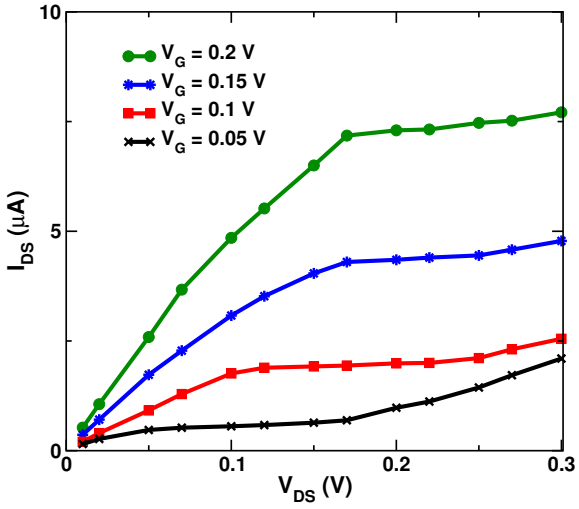

Figure 3.18: Output characteristics of the simulated GNR-FET with SB contacts and LDD extension. The doping profile is exponential and different for each curve. For V_G equal to 0.05 V (black curve), it goes from $3 \cdot 10^{-5}$ to $2.5 \cdot 10^{-4}$, for V_G equal to 0.1 V (red curve) from $5 \cdot 10^{-5}$ to $1 \cdot 10^{-4}$ and for V_G equal to 0.15 V (blue curve) from $1 \cdot 10^{-4}$ to $2 \cdot 10^{-4}$. The characteristic for V_G= 0.2 V is as illustrated in fig. 3.15.

effects on total current and is not the best strategy to be used.

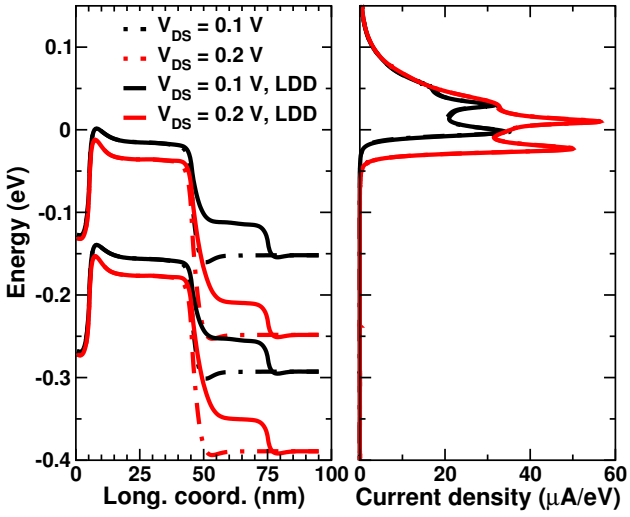

Figure 3.19: Band diagrams (left) and current spectra (right) calculated for the two bias points V_G = 0.1 V, V_{DS} = 0.1 V (black lines) and V_G = 0.1 V, V_{DS} = 0.2 V (red lines) for the device in fig. 3.1 with the parameters as in fig. 3.13.

3.4 Generalization of the study

The results presented in previous sections suggest that to maximise 10 to 15 nm-wide GNR-FETs performance, two main issues induced by BTBT have to be specifically addressed: the first is contrasting the reduction of the potential energy barrier due to holes pile-up in the channel, while the second is limiting the hole current flowing directly from the source to the drain. To handle the energy barrier lowering, the most suitable strategy is reducing the EOT, as thinner oxides provide better electrostatic control over the channel. That can be conveniently obtained by using high-k oxides as dielectric, as proposed in sec.3.2.2. Therefore, the source region is no more required to be kept very short to favour the escape of holes from the channel. Thus, semi-infinite doped leads could be assumed, but in this study metal contacts are considered, in order to simulate more realistic devices and achieve better accuracy of the results. To get rid of the current reduction induced by metal contacts because of the Schottky barriers between the source/drain regions and the channel, those regions have been heavily doped, thus obtaining almost ohmic contacts. In fact, heavy doping leads to strong bending of the energy bands, resulting in SB barriers that are easily tunneled through by carriers. For the simulation study in this section, symmetric doping equal to $2 \cdot 10^{-3}$ dopant atoms/carbon atoms has been taken in the source and drain regions. Given that it is not necessary that the source region is short enough to be easily tunneled through by holes, the length of that region can become a design parameter, that could be optimized to achieve better performance. Therefore, it is important to define the optimum length L_S, in consideration of a reasonable tradeoff between occupancy and capability of limiting the hole current flowing directly from the source to the drain. In fig.3.20, the band diagrams and current spectra are illustrated for the bias points V_G = 0.2 V, V_{DS} = 0.2 V(left panel) and 0.35 V (right panel) for L_S ranging from 5 to 20 nm. While for V_G = 0.2 V, V_{DS} = 0.2 V the picture does not change significantly, increasing the length of the source region results in a strong limitation of the BTBT at higher V_{DS}, as evident in the plot in the right panel. As different lengths of the source region induce different location of the resonant states, peaks in the current density plots slightly differ from case to case, but the total current generally decreases as source length increases. With regard to the bias points shown in fig.3.20, total current decreases from 11.1 μA for L_S= 5 nm to 10.7 μA for L_S= 20 nm at V_G = 0.2 V, V_{DS} = 0.2 V, while from 15.4 μA to 13.9 μA at V_G = 0.2 V, V_{DS} = 0.35 V. Therefore, for the analysis performed in this section, L_S is set to 20 nm. With the purpose of providing reasonable trade-off between speed performance and limitation of direct source-to-drain tunneling phenomena, even L_D and gate length are set to 20 nm as well. The turnon and output character-

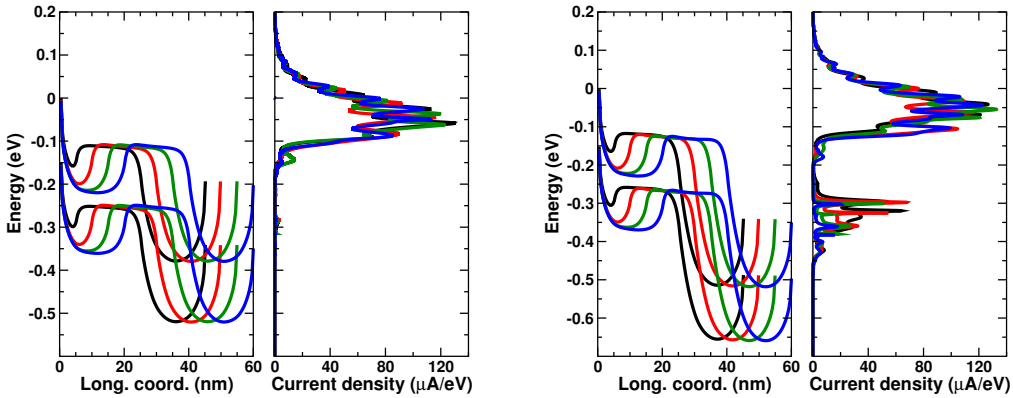

Figure 3.20: Band diagrams (left) and current spectra (right) calculated for the bias points V_G = 0.2 V, V_{DS} = 0.2 V(left panel) and 0.35 V (right panel) for different L_S = 5/10/15/20 nm. The gate and drain lengths have been taken equal to 20 nm. Doping in the source and drain regions has been set to $2 \cdot 10^{-3}$ dopant atoms/carbon atoms.

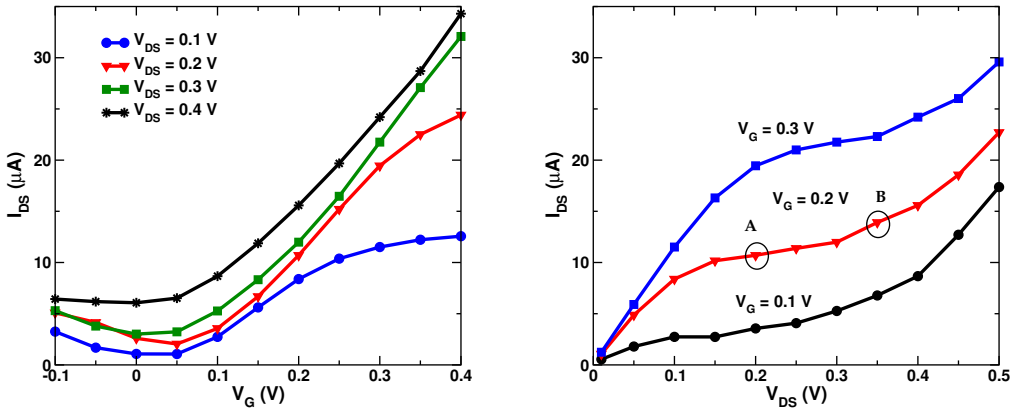

Figure 3.21: Turn-on (left panel) and output characteristics (right panel) of the simulated GNR-FET with L_S=L_G=L_D= 20 nm, HfO_2 thickness= 2 nm, GNR width= 10 nm. Labels A (at V_G= 0.2 V, V_{DS}= 0.2 V) and B (at V_G= 0.2 V, V_{DS}= 0.35 V) mark the bias points relative to the plots of figs. 3.22 and 3.25.

istics of the devices under investigation are illustrated in fig. 3.21 in the left and right panel, respectively. Replacing silicon oxide with a high-k dielectric and increasing the length of source region are certainly a good strategy, but could not be sufficient to overcome the current rise at high V_{DS} due to the small band-gap. The most suitable approach is therefore combining that with other techniques, which are effective in contrasting the current flow of high-energy holes from source to drain. The introduction of a lightly-doped drain extension, as proposed in sec. 3.3.1, could

work, but poses heavy technological challenges, like the accurate control of the doping profile in the LDD region and the choice of the bias region to be optimized. In the perspective of suggesting easily manufacturable solutions, here the introduction of an electrostatically doped region is suggested, as proposed for carbon nanotubes (CNTs) in [53, 54] and graphene in [6, 55]. Thus, the simulated structure relative to fig. 3.21, is modified by the introduction of an additional set of top and bottom gates. In this study, 5 nm has been taken as the distance between the gates, as it is on one hand it would suffice to prevent technological problems and on the other it is short enough to keep good electrostatic control over the channel between them. For the same purpose of obtaining good electrostatic control over the channel, also the gate length of the second set of gates is taken equal to 20 nm. Realizing the second set of gates by choosing a different metal (with the appropriate work function difference with respect to the first gate) will result in a modification of the energy band profiles of the structure, that could lead to a significant suppression of BTBT, as it will be discussed further. The comparison of the band diagrams and current spectra for the bias points V_G = 0.2 V, V_{DS} = 0.2 and V_G = 0.2 V, V_{DS} = 0.35 V for the proposed architectures are illustrated in fig. 3.22, where the solid lines refer to the device with one double-gate and the dashed lines to the two double-gate device with work function difference between metals used for the gates equal to 0.1 eV and length of the second gate set equal to the one of the first gate, namely 20 nm. The use of hafnium oxide is con-

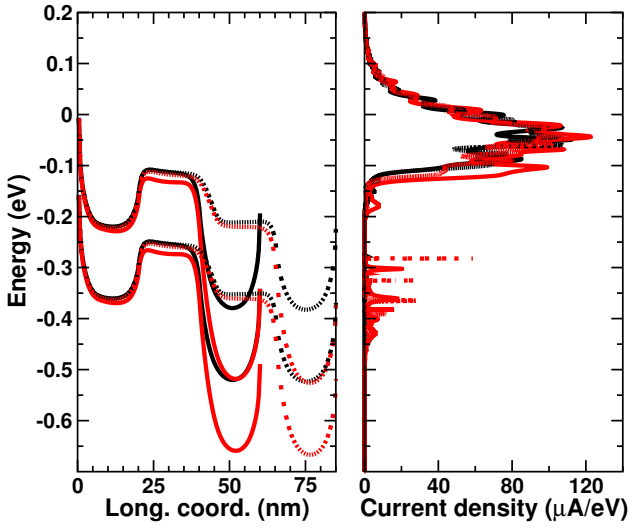

Figure 3.22: Band diagrams (left) and current spectra (right) calculated for the bias points V_G = 0.2 V, V_{DS} = 0.2 V(black lines) and 0.35 V (red lines). The simulated structure has one gate (solid lines) or two sets of gates with work function difference equal to 0.1 eV (dashed lines).

firmed to be a good strategy to limit the energy barrier lowering caused by the strong BTBT. In fact, the current dependance on drain voltage is quite weak. However, the band diagrams of the one-gate device still present a mild potential energy barrier lowering as the drain voltage increases. That results in a high peak of the current density around $E = -0.1$eV and therefore a slight current increase. Within this scenario, the introduction of a second set of top and bottom gates with the appropriate work function, leads to a noticeable increase of the tunneling distance, thus to a strong reduction of the BTBT. In this case, depicted in dashed lines in 3.22, apparently no difference is observable between the energy barrier heights for the band profiles at the two chosen bias points. Also the peak of the current density at $E = -0.1$ eV is significantly reduced. In addition to that, in the energy range between $E = -0.3$ eV and $E = -0.5$ eV, the hole current is remarkably limited and the current density plot shows only a few sharp peaks, ascribed to resonant states in the valence band. With the purpose of further optimization, the length of the second set of gates has been taken as a design parameter. Fig.3.23 illustrates the band diagrams and current spectra when the length of the second set of gates L_{G2} has been reduced from 20 nm to 10 nm. As expected, the longer is the second gate contact, the stronger is the BTBT suppression. Although the current spectra show only minor difference, the total current is remarkably reduced from 12.8 μA to 11.7 μA. Further lengthening the second gate length is not advisable, since it would negatively impact the speed performance of the

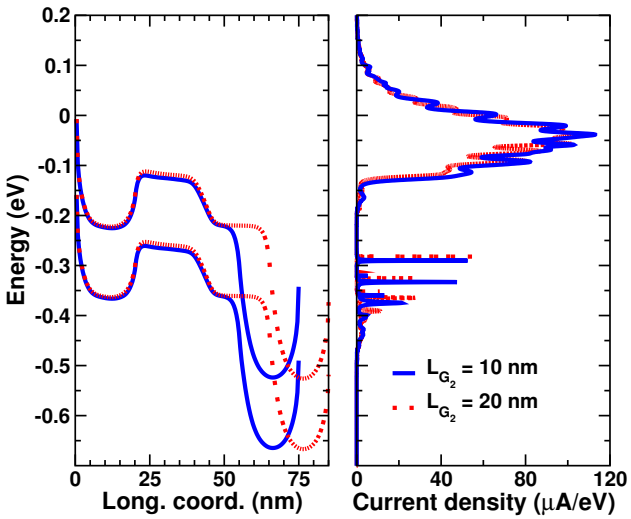

Figure 3.23: Band diagrams (left) and current spectra (right) calculated for the bias point V_G = 0.2 V, V_{DS} = 0.35 V for an architecture with two sets of gates for different lengths of the second gate contact. The work function difference has been taken equal to 0.1 eV.

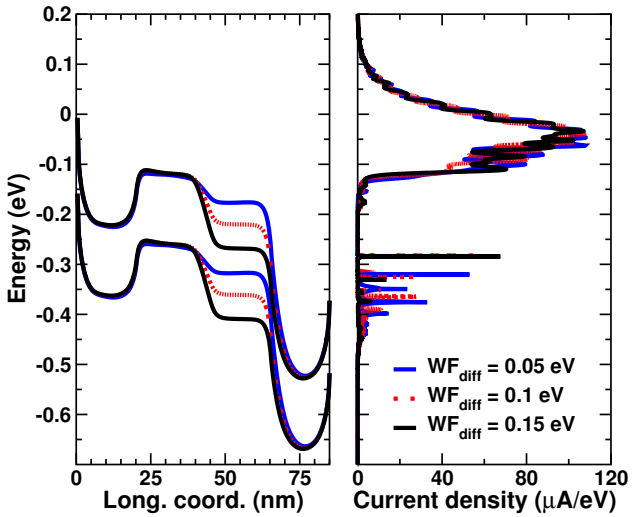

Figure 3.24: Band diagrams (left) and current spectra (right) calculated for the two-gate structure at the bias point V_G = 0.2 V, V_{DS} = 0.35 V. The work function difference (WF_{diff}) between the first and second set of gates is 0.05 eV (solid blue lines), 0.1 eV (dashed red lines) or 0.15 eV (solid black lines).

device. Therefore, from now on, the study will be limited to additional sets of contacts with length equal to 20 nm.

The effectiveness of the introduction of the second set of gates in contrasting BTBT phenomena is clearly related to the work function difference between the metals they are made of. Figure 3.24 illustrates the band diagrams and current spectra at the bias V_G = 0.2 V, V_{DS} = 0.35 V for the two double-gate device discussed above, where the work function difference between the first and second set of gates is set to 0.05 eV, 0.1 eV or 0.15 eV. As a result, the electron current spectra present only minor difference, while the peaks in the hole current density are apparently reduced when using metals with higher work function difference. This results in a decrease of the total current from 12.6 μA when using two metals with work function difference equal to 0.05 eV to about 11.7 μA when using metals with work function difference equal to 0.1 or 0.15 eV. This suggests that the most suitable choice of metal sets that fully exploits the potential of the multi-gate architecture is the one for which the lowest conduction band in the portion of the channel controlled by the second gate is positioned between the mid-gap and the top of the valence band relative to that part of the channel controlled by the first gate. Thus, in the study proposed in this section, the reference work function difference will be taken equal to 0.1 eV. Though the proposed two-gate approach significantly limits the phenomena related to the small band-gap and extends the saturation region

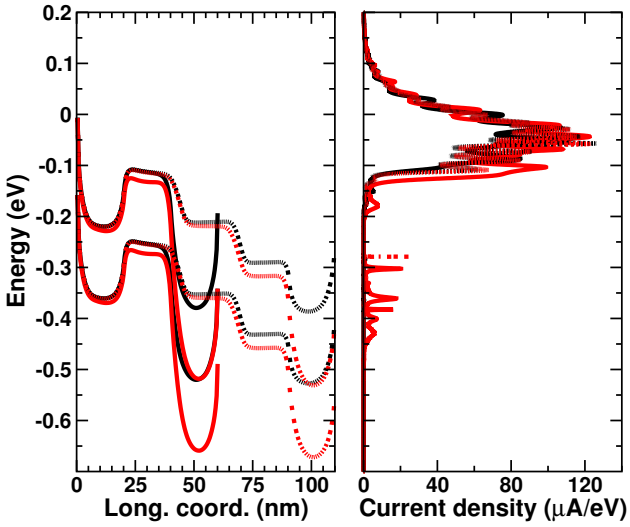

Figure 3.25: Band diagrams (left) and current spectra (right) calculated for the bias points as in fig.3.22. The simulated structure has one gate (solid lines) or three sets of gates with work function difference equal to 0.1 eV (dashed lines).

of the output characteristics, one could argue that it could be not sufficient to achieve current saturating behaviour within the full bias range. Therefore, the introduction of an additional set of top/bottom gates could be appropriate if the device is going to be biased at high drain voltages, for which the two double-gate structure limits, but not suppresses completely, the hole current. In fig.3.25 are represented the band diagrams and current spectra of the one double-gate and the three double-gate structure at the same bias point as in fig.3.22. The work function difference between the first and second set of gates and between the second and third sets is 0.1 eV, thus the total difference between the first and third sets is 0.2 eV. As expected, the high-energy electron current spectra almost overlap and the peak at E=-0.1 eV is significantly reduced. Unlike the two double-gate case presented above, for the three double-gate structure the hole current spectrum is almost negligible and only one minor peak is observable. This confirms the great potential of the multi-gate approach in extending the bias range for which current saturating behaviour is achieved, thus leading to high voltage-gain and wider operation region. The output characteristics at three different gate voltages, namely 0.1 V, 0.2 V and 0.3 V (respectively in black, red and blue lines) for the one, two and three double-gate structures are depicted in fig.3.26. The benefit of the addition of second and third sets of gates is evident, especially at high drain voltage. As mentioned before, the inclusion of one additional set of gates significantly extends the saturation region of about 0.1 V for any gate voltage. While

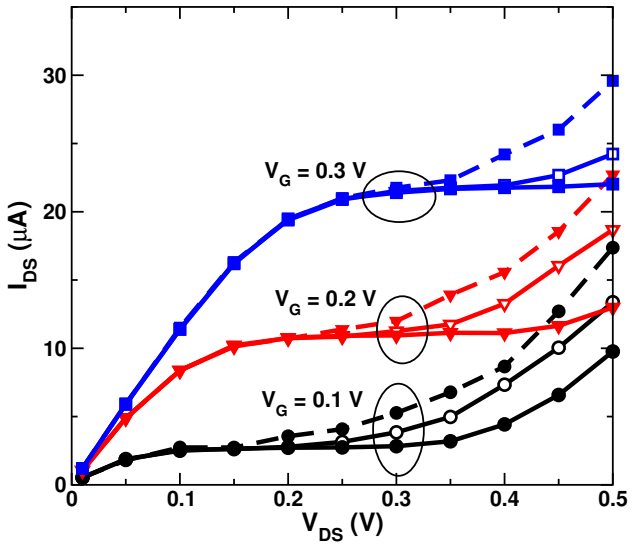

Figure 3.26: Output characteristics for the one double-gate (dashed lines, closed symbols), two double-gate (solid lines, open symbols) and three double-gate (solid lines, closed symbols) structures. The GNR-width is 10 nm and all the gates are 20 nm long. The work function difference between two adjacent gates has been set to 0.1 eV. The plots relative to the one double-gate architecture, previously shown in the right panel of fig. 3.21, have been reported here to facilitate comparison.

further addition of another set of gates on one hand results in extending the operation region up to $V_{DS} = 0.4$ V, on the other hand it determines an increase of the voltage gain but also a decrease of the cutoff frequency. The designer should keep in mind the trade-off between saturation and transit time when determining the most suitable number of gates for the device, depending on the application it is designed for. The turn-on characteristics for the three double-gate architecture, which should be compared with the left panel of fig. 3.21, are shown in fig.3.27. Clearly, the multi-gate structure is less sensitive than the one double-gate topology to variations of V_{DS} at high V_G, for which the current is mainly determined by the flow of electrons over the potential energy barrier. In fact the multi-gate approach strongly reduces the BTBT, thus minimizing the hole current, as shown in fig. 3.25. With respect to the on/off current ratio, the improvement achieved by using the multi-gate architecture at drain voltage up to 0.3 V is apparent. As an example, at V_{DS} equal to 0.3 V the on/off current ratio (calculated as $I_{DS(V_G=0.4V)}/I_{DS(V_G=0V)}$) for the three double-gate structure is about 50, versus 10 calculated for the one double-gate device. The remarkable increase of that figure of merit is related to the significant limitation of the current flowing directly from the source to the drain at

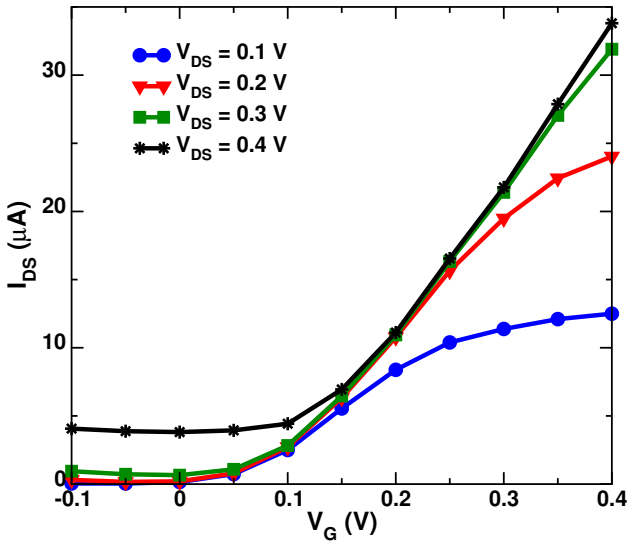

Figure 3.27: Turn-on characteristics for three double-gate structure for V_{DS} from 0.1 V to 0.4 V. Simulation parameters are the same as in fig. 3.26.

low V_G, as illustrated in fig. 3.28 for the bias point V_G= 0 V, V_{DS}= 0.3 V. The two main peaks of the current spectrum are almost suppressed and the BTBT at the drain end of the channel has only minor impact on the potential energy barrier height. This results in an off current decrease by a factor of about 5. However, at low gate voltage and very high drain voltage there is no observable difference between off currents for the two proposed architectures. This is mainly related to the fact that current at such bias is dominated by high-energy holes tunneling from the valence band in the source to the conduction band in the drain. Figure 3.29 depicts band diagrams and current spectra for the three double-gate architecture at low V_G equal to 0 V and different drain voltages, namely 0.1 V, 0.3 V and 0.4 V. It is observed that when the conduction band in the drain begins to occupy the same energy levels of the valence band in the source, the current ascribable to direct flow from source to drain starts increasing rapidly and at V_{DS} equal to 0.4 V it becomes significant. At that bias point, clearly the three double-gate approach does not provide any advantage compared to the standard one double-gate architecture since any energy band-gap is present for carriers on their way to the drain. In fact, current levels in those two cases are comparable, although the current is slightly smaller for the three double-gate device. To complete the study, it is important to measure at what extent the multi-gate structure is more suitable than the standard one double-gate architecture for analog applications, by comparing their small-signal parameters. As reference, the bias point for which the transconductance g_m is maximum has been chosen since it is the one for which both A_V and f_T are maximized. Thus, for V_{DS} equal to

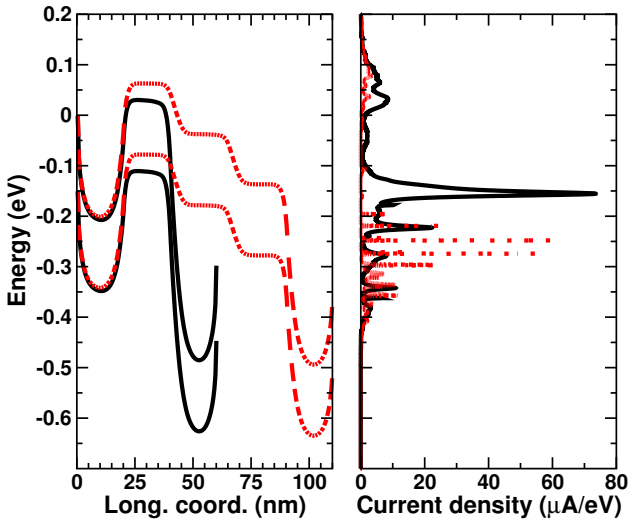

Figure 3.28: Band diagrams (left) and current spectra (right) calculated for the bias point V_G= 0, V_{DS}= 0.3 V for the one double-gate structure (solid black lines) and three double-gate structure (dashed red lines).

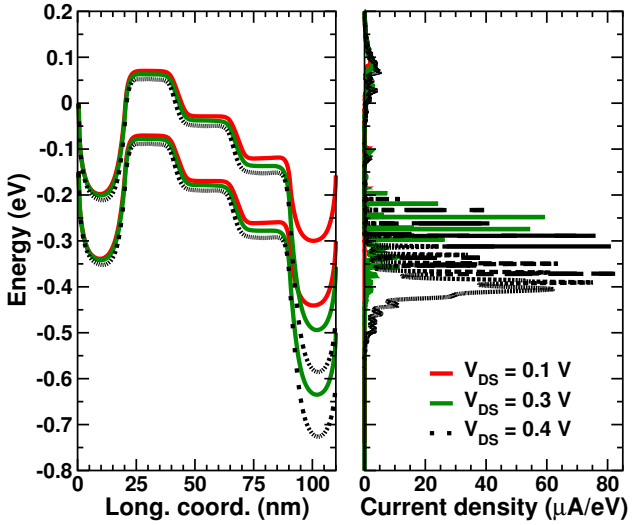

Figure 3.29: Band diagrams (left) and current spectra (right) calculated at V_G= 0 for different drain voltages, namely V_{DS}= 0.1 V (solid red lines), V_{DS}= 0.3 V (solid green lines), V_{DS}= 0.4 V (dashed black lines) for three double-gate structure.

0.35 V, g_m has been calculated by numerical approximation with a 3 mV-voltage step, as in fig. 3.6. The plots of g_m with and without the inclusion of acoustic and optical phonons are shown in fig. 3.30. The highest values of g_m are in the range 0.3 V<V_G<0.4 V. Specifically, in ballistic conditions the transconductance shows a slightly oscillating behaviour with

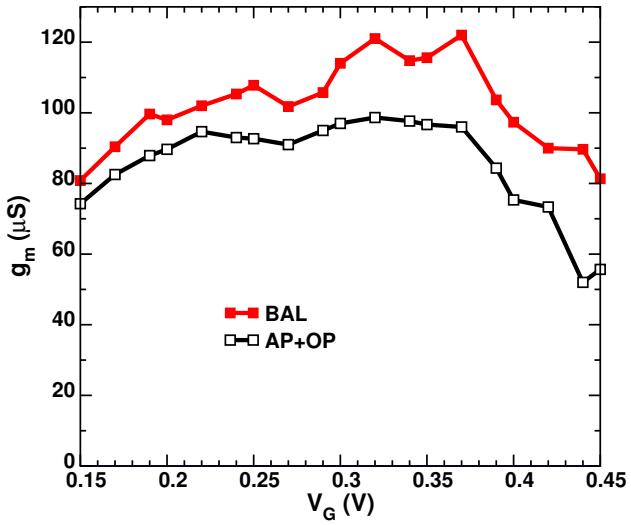

Figure 3.30: Transconductance vs. gate voltage at V_{DS}= 0.35 V for the three double-gate structure in ballistic conditions (red line, closed symbols) or with the inclusion of acoustic and optical phonons (black line, open symbols). The model used for the treatment of phonons is described in chapter 1.

respect to the gate voltage, mainly due to the presence of resonant states in the valence band, with two peaks at V_G= 0.32 V and V_G= 0.37 V. On the contrary, the broadening of the density of states, induced by the inclusion of phonons in the picture, induces a smoother profile of the g_m, even though the current levels and thus g_m itself are lower than in ballistic case. As observed in ballistic conditions, the top g_m is obtained at V_G= 0.32 V, therefore the bias point V_G= 0.32 V, V_{DS}= 0.35 V has been set as reference for the evaluation of performance. Table 3.3 summarizes the small-signal figures of merit for all the architectures (one/two/three double-gate) discussed above. As expected, further addition of gates significantly reduces the BTBT, thus resulting in a noticeable reduction of the output conductance. Therefore, a remarkable increase of the voltage-gain is achieved, while the increased gate capacitance has adverse impact on the cutoff frequency, although f_T shows high performance (almost 700 GHz) even in the worst scenario. Therefore, depending on A_V and f_T required for the desidered application, the designer could take the number of gates as an additional design parameter. The inclusion of elastic and optical phonons in the model of the three double-gate devices impacts the voltage-gain more than the cutoff frequency, causing A_V to decrease from 39.8 to 12.9 and f_T from 0.68 to 0.55, thus keeping quite good analog performance. As the GNR width increases, the limitations induced by the small band-gap become more severe. To generalize the considerations above, the study

N	$g_m\ [\mu S]$	$g_d\ [\mu S]$	$C_{gs}\ [aF]$	A_v	$f_T\ [THz]$
1	103.5	23.5	8.3	4.4	1.98
2	111	12.5	16	9.2	1.1
3	121	3	28.4	39.8	0.68

Table 3.3: Summary of the main results obtained at the reference bias (V_G= 0.32 V, V_{DS}= 0.35 V) for 10 nm-wide GNR-FETs using hafnium oxide as dielectric. Different architectures, namely one/two/three double-gate structures with L_G equal to 20 nm, are proposed. The number of gates (1, 2 or 3) is reported in the first column. For multi-gate devices, a work function difference between the first and second (and second/third) set(s) equal to 0.1 eV has been used. Simulations have been performed in ballistic conditions.

has been extended to wider GNR-FETs. Therefore, 15 nm-wide graphene nanoribbons have been taken as channel material, using the same parameters as in table 3.3. Fig.**??** illustrates the turnon and output characteristics of the device. Altough the plots in fig.3.31, compared with figs.3.27 and 3.26 concerning the 10 nm-wide device, confirm that an increase of the GNR width results in a degradation of performance in terms of both the current on/off ratio and saturation at high V_{DS}, the introduction of the multi-gate approach remarkably improves performance. For example, at V_{DS}= 0.1 V the ratio between the current in ON-state (calculated at V_G= 0.4 V) and current in OFF-state (calculated at V_G= 0 V) moves from about 4 for the one-double gate architecture to 14 for the 3-double gate topology. At V_{DS}= 0.2 V and V_{DS}= 0.3 V the I_{ON}/I_{OFF} ratio, calculated at the same bias points, has been improved from about 5 to 30, while at V_{DS}= 0.3 V from almost 5 to 16. Even though the improvement of performance is apparent, wide GNR-FETs are confirmed not suitable for logic applications. The small-signal parameters obtained from the simulation study are collected in table 3.4, where the first column shows the type of achitecture proposed, as in table 3.3, and the column (WF_{diff}) reports the work function difference between two adjacent gates, that is taken as an additional parameter of the investigation. As discussed above, the 15 nm-wide GNR-FETs show higher current with respect to the 10 nm-wide devices, thus resulting in higher g_m at a given bias. Therefore one can assert that the voltage-gain for the standard one double-gate architecture is definitely larger than in the previous case. Unlike observed for devices that use narrower nanoribbons, the two double-gate architecture would not be effective in contrasting BTBT, for any work function difference between gates. Even for the inclusion of the third set of gates with work function difference equal to 0.1 eV the performance would not improve. On the contrary, with the inclusion of the third set of gates with low work function differ-

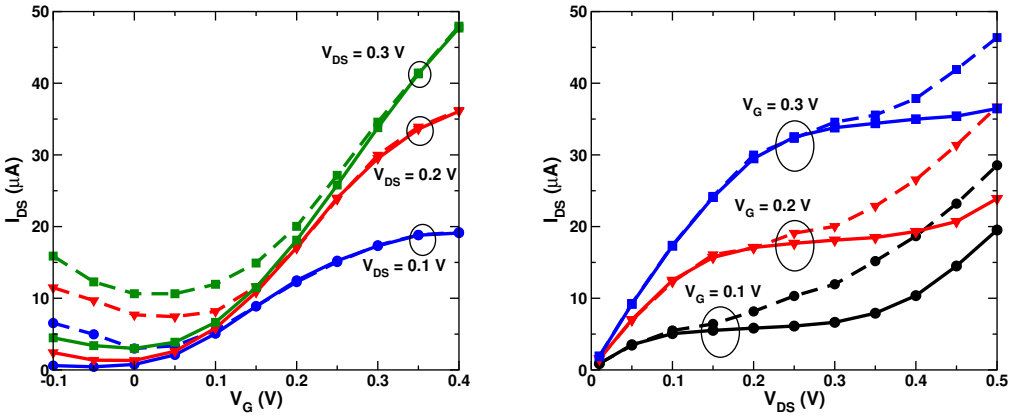

Figure 3.31: Turn-on (left panel) and output characteristics (right panel) of the simulated 15 nm-wide GNR-FET, with the same parameters listed in fig.3.26. Calculated data for both the one-double gate (dashed lines) and three-double gate (solid lines) architecture are reported.

N	$WF_{diff}\ [eV]$	$g_m\ [\mu S]$	$g_d\ [\mu S]$	$C_{gs}\ [aF]$	A_v	$f_T\ [THz]$
1		161	15	14.1	10.7	1.82
2	0.1	172.7	20.3	24.1	8.5	1.14
	0.05	156.9	33.3	25.3	4.7	0.99
3	0.1	181	25.7	38.3	7.1	0.75
	0.05	163	10.3	29.7	15.8	0.87

Table 3.4: Summary of the main results obtained for 15 nm-wide GNR-FETs. Simulation parameters and reference bias are the same as in tab. 3.3.

ence one achieves the expected advantage. In fact, g_m is very close to the one obtained with the one gate architecture, while g_d is definitely smaller, thus resulting in higher A_V. As discussed before, the inclusion of additional gates adversely impacts the cutoff frequency, but it still maintains good performance, above 800 GHz. Even though good performance have generally been achieved, the key point regarding 15 nm-wide GNR-FETs is that the smaller band-gap requires better accuracy in determining the most suitable sets of metals to be used, since varying the difference between the work functions from 50 meV to 100 meV causes quite a degradation in performance, as reported in table 3.4. The explanation of this behaviour is clearly that the position of the conduction and valence bands in the different sections of the channel when the work function difference is higher is not the one that minimizes BTBT, as illustrated in fig. 3.32.

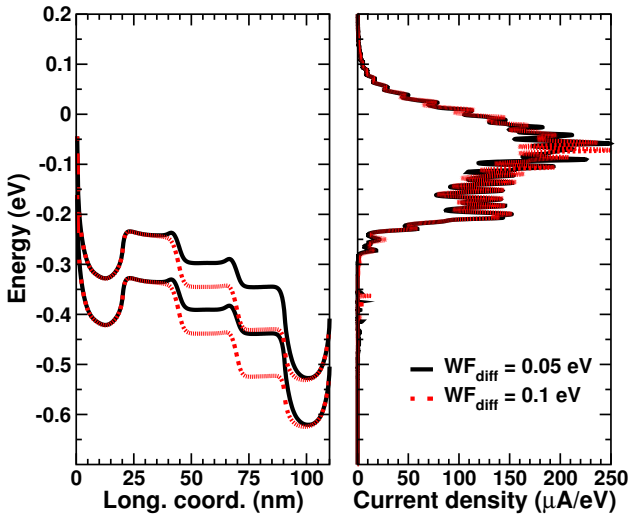

Figure 3.32: Band diagrams (left) and current spectra (right) calculated for a 15 nm-wide GNR-FET with three double-gate architecture at the bias point V_G= 0.32 V, V_{DS}= 0.35 V.

3.5 Summary

In this chapter the high-frequency analog performance of 10-15 nm-wide GNR-FETs has been investigated by means of simulations based on a full-quantum atomistic model. Ballistic conditions have been considered, although in some cases the inclusion of acoustic and optical phonon scattering has been taken into account. The small band-gap typical of these GNRs induces band-to-band-tunneling phenomena, which increase with the drain voltage, thus preventing current from saturating and considerably affecting the small-signal voltage-gain. Besides confirming the excellent cut-off frequency potential for these devices, this study highlights some limitations in the maximum achievable voltage-gain. A_V is limited by effects related to the small band-gap and not by short-channel effects, as opposed to standard MOSFETs. Thus, the main concern is that the analog designer can not trade-off gain and speed by choosing a non-minimum gate length. Design criteria to contrast the effect of band-to-band tunneling for GNR-FETs with both semi-infinite leads and metal contacts have been suggested. In particular, for the first devices, asymmetrical doping in the source and drain regions and use of high-k oxide, while for the second the introduction of a LDD region and careful choice of the length of the source region have been shown to be the best techniques to maximize performance. A multi-gate approach, combined with the use of high-k oxide and careful determination of the length of the source region, has been suggested as the best strategy to achieve current saturation and ex-

tending the operation region for the devices under investigation. For 10 nm-wide GNR-FETs, very good analog performance (A_V almost equal to 40 and f_T up to 700 GHz) has been obtained. The designer could take the number of gates as an additional parameter, even though keeping in mind that, the wider the GNR, the more gates could be necessary to satisfy the performance requirements and the more accurate the selection of the appropriate set of gate metals should be.

Conclusions

In this work, performance of graphene nanoribbon-FETs with width ranging from 1 to 15 nm has been extensively investigated through numerical simulations. A code has been developed for the simulation of GNR-FETs: a full-quantum approach has been adopted, through the use of a TB NEGF formulation, including edge roughness and scattering with elastic and optical phonons. Given the high computational cost of this model, numerical techniques have been proposed to speed-up calculations without lacking in numerical accuracy.
In the first part of this work, very narrow GNR-FETs with both smooth and rough edges have been studied. It has been found that quantum localization effects due to edge roughness become important, so that it is not possible to define an ER limited mobility. In this regime, the inclusion of acoustic phonon scattering results in a current increase rather than decrease, due to phase breaking, which partially recovers the ohmic regime. The resulting total mobility extracted from a long channel GNR-FETs biased in the ON-state and localization length are close to data available in literature. Although the inclusion of elastic phonon scattering partially recovers the diffusive law and reduces the variability, for narrow GNR-FETs edge roughness remains the main current limiting effect, degrading mobility much more severely than elastic phonon scattering.
With respect to GNR-FETs wider than 10 nm, it has been shown that the performance is mainly limited by the effects of band-to-band tunneling. Thus, design criteria to contrast BTBT in GNR-FETs with both semi-infinite leads and metal contacts have been suggested. A multi-gate approach, combined with the use of high-k oxide and careful determination of the length of the source region, has been advised as the best strategy to achieve current saturation and extending the operation region for the devices under investigation. For 10 nm-wide GNR-FETs, very good analog performance (A_V almost equal to 40 and f_T up to 700 GHz) has been obtained. Therefore, 10-to-15 nm-wide GNR-FETs are promising candidates for analog applications that require high-frequency operation but are not very demanding in voltage gain, such as low-noise amplifiers (LNAs).

Bibliography

[1] Y. Taur et al., "Fundamentals of Modern VLSI Devices", Second ed., Cambridge University Press, 2009.

[2] K. S. Novoselov et al., "Electric Field Effect in Atomically Thin Carbon Films", *Science*, vol. 306, n. 5696, pp. 666-669, 2004.

[3] A. K. Geim, "Graphene: Status and Prospects", *Science*, vol. 324, n. 5934, pp. 1530-1534, 2009.

[4] K. S. Novoselov et al., "Two-dimensional gas of massless Dirac fermions in graphene", *Nature*, vol. 438, pp. 197-200, 2005.

[5] Y. B. Zhang et al., "Experimental observation of the quantum Hall effect and Berry's phase in graphene", *Nature*, vol. 438, pp. 201-204, 2005.

[6] F. Wang et al., "Gate-variable optical transitions in graphene", *Science*, vol. 320, pp. 306-309, 2008.

[7] J. H. Chen et al., "Intrinsic and extrinsic performance limits of graphene devices on SiO2", *Nature Nanotechnol.*, vol. 3, pp. 206-209, 2008.

[8] J. H. Seol et al., "Two-dimensional phonon transport in supported graphene", *Science*, vol. 328, pp. 213-216, 2010.

[9] Y.-M. Lin et al., "100 GHz transistors from wafer-scale epitaxial graphene", *Science*, vol. 327, 662, 2010.

[10] P. Avouris et al., "Carbon-based electronics", *Nature Nanotechnol.*, vol. 2, pp. 605-615, 2007.

[11] A. H. Castro Neto et al, "The electronic properties of graphene", *Reviews of Modern Physics*, vol. 81, pp. 109-54, 2009.

[12] G. Fiori et al., "Simulation of Graphene Nanoribbon Field-Effect Transistors", *IEEE Electron Device Lett.*, vol. 28, n. 8, pp. 760-762, 2007.

[13] R. Saito et al., "Electronic structure of chiral graphene tubules", *Appl. Phys. Lett.*, vol. 60, pp. 22042206, 1992.

[14] T. W. Odom et al., "Atomic structure and electronic properties of single-walled carbon nanotubes", *Nature*, vol. 391, pp. 62-64, 1998.

[15] http://www.mrsec.umn.edu/Research/Seed2010$_{K}oester$.php.

[16] M. Y. Han et al., "Energy band-gap engineering og graphene nanoribbons", *Phys. Rev. Lett.*, vol. 98, p. 206805, 2008.

[17] Y.-W. Son et al., "Energy gaps in graphene nanoribbons", *Phys. Rev. Lett.*, vol. 97, , p. 216803, 2006.

[18] S. Datta, "Electronic Transport in Mesoscopic Systems", *Cambridge University Press*, 1997.

[19] R.Lake et al., "Single and multiband modeling of quantum electron transport through layered semiconductor devices", *J. Appl. Phys.*, vol. 81, n. 12, p. 7845, 1997.

[20] M. P. Lopez Sancho et al., "Highly convergent schemes for the calculation of bulk and surface green functions", *J. Phys. F: Met. Phys.*, vol. 15, p. 851, 1985.

[21] R. Grassi et al., "Mode Space Approach for Tight-Binding Transport Simulation in Graphene Nanoribbon FETs", *IEEE Trans. on Nanotech.*, vol. 10, n.2, pp. 371-378, 2011.

[22] R. Grassi, Ph.D. thesis, 2011.

[23] L. P. Kadanoff et al., "Quantum Statistical Mechanics", *Benjamin*, 1962.

[24] M. Frey et al., "Simulation of intravalley acoustic phonon scattering in silicon nanowires", *Proc. Essderc 2008*, p. 258, 2008.

[25] D. Finkenstadt et al., "From Graphene to Graphite: A General Tight-binding Approach for Nanoribbon Carrier Transport", *Phys. Rev. B*, n. 76, p. 121405, 2007.

[26] M. Frey et al., "Boundary Conditions for Incoherent Quantum Transport", *Proc. IWCE 2009*, p. 17, 2009.

[27] J. H. Chen et al., "Intrinsic and Extrinsic Performance Limits of Graphene Devices on SiO2", *ArXiv*.

[28] Y. Yoon et al., "Role of phonon scattering in graphene nanoribbon transistors: Nonequilibrium Greens function method with real space approach", *Appl. Phys. Lett.*, vol. 98, p. 203503, 2011.

[29] X. Wang et al., "Room-temperature all-semiconducting sub-10-nm graphene nanoribbon field-effect transistors", *Phys. Rev. Lett.*, vol. 100, p. 206803, 2008.

[30] B. Obradovic et al., "Analysis of graphene nanoribbons as a channel material for field-effect transistors", *Appl. Phys. Lett.*, vol. 88, p. 142102, 2006.

[31] T. Fang et al.,"Mobility in semiconducting graphene nanoribbons: Phonon, impurity, and edge roughness scattering", *Phys. Rev. B*, vol. 78, p. 205403, 2008.

[32] M. Bresciani et al., "A Better Understanding of the Low-Field Mobility in Graphene Nano-ribbons", *Proc. ESSDERC 2009*, pp. 480-483, 2009.

[33] Y. Yoon et al., "Effect of edge roughness in graphene nanoribbon transistors", *Appl. Phys. Lett.*, vol. 91, pp. 073103, 2007.

[34] R. Grassi et al., "An investigation of performance limits of conventional and tunneling graphene-based transistors", *J. Comput. Electron.*, vol. 8, pp. 441-450, 2009.

[35] A. Betti et al., "Strong mobility degradation in ideal graphene nanoribbons due to phonon scattering", *Appl. Phys. Lett.*, vol. 98, p. 212111, 2011.

[36] Y. Yoon et al., "Effect of edge roughness in graphene nanoribbon transistors", *Appl. Phys. Lett.*, vol. 91, p. 073103, 2007.

[37] A. Svizhenko et al., "Effect of growth orientation and surface roughness on electron transport in silicon nanowires", *Phys. Rev. B*, vol. 75, p. 125417, 2007.

[38] A. Cresti et al., "Range and correlation effects in edge disordered graphene nanoribbons", *New Journal of Physics*, vol. 11, p. 095004, 2009.

[39] *The International Technology Roadmap for Semiconductors* http://www.itrs.net/Links/2011ITRS/Home2011.htm.

[40] L. Jiao et al., "Narrow graphene nanoribbons from carbon nanotubes", *Nature*, vol. 458, pp. 877-880, 2009.

[41] X. Wang et al., "Graphene nanoribbons with smooth edges behave as quantum wires", *Nature Nanotech.*, vol. 6 , pp. 563-567 , 2011.

[42] A. D. Liao et al., "Thermally Limited Current Carrying Ability of Graphene Nanoribbons", *Phys. Rev. Lett.*, vol. 106, p. 256801 , 2011.

[43] X. Li et al., "Chemically Derived, Ultrasmooth Graphene Nanoribbon Semiconductors", *Science*, vol. 319 , pp. 1229-1232, 2008.

[44] L. Xie et al., "Graphene Nanoribbons from Unzipped Carbon Nanotubes: Atomic Structures, Raman Spectroscopy, and Electrical Properties", *JACS*, vol. 133, pp. 10394-10397, 2011.

[45] Y.-M. Lin et al., "Development of graphene FETs for high frequency electronics", *Tech. Dig. IEDM*, pp.237-240, 2009.

[46] I. Meric et al., "RF performance of top-gated, zero band-gap graphene field-effect transistors", *Tech. Dig. IEDM*, pp. 1-4, 2008.

[47] Y.-M. Lin et al., "Operation of graphene transistors at gigahertz frequencies", *NanoLett.*, vol. 9, pp. 422-426, 2009.

[48] L. Liao et al., "High-speed graphene transistors with a self-aligned nanowire gate", *Nature*, 467, pp.305-308, 2010.

[49] F. Schwierz, "Graphene transistors", *Nature Nanotechnol.*, vol.5, pp. 487-496, 2010.

[50] J. Lee et al., "RF performance of pre-patterned locally-embedded-back-gate graphene device", *Tech. Dig. IEDM*, pp. 568-571, 2010.

[51] J. S. Moon et al., "Epitaxial graphene RF field-effect transistors on Si-face 6H-SiC substrates", *IEEE Electron Device Lett.*, vol. 30, pp. 650-652, 2009.

[52] S. Das et al., "On the Importance of Bandgap Formation in Graphene for Analog Device Applications", *IEEE Trans. on Nanotech.*, vol. 10, no. 5, pp. 1093-1098, 2011.

[53] J. Knoch et al., "Tunneling phenomena in carbon nanotube field-effect transistors", *Phys. Stat. Sol. A*, vol. 205, no. 4, pp. 679-694, 2008.

[54] J. Appenzeller et al., "Band-to-band tunneling in carbon nanotube field-effect transistors", *Phys. Rev. Lett.*, vol. 93, no. 19, pp. 196805-1-196805-4, 2004.

[55] Y. Zhang et al., "Direct observation of a widely tunable bandgap in bilayer graphene", *Nature*, vol. 459, pp. 820-823, 2009.